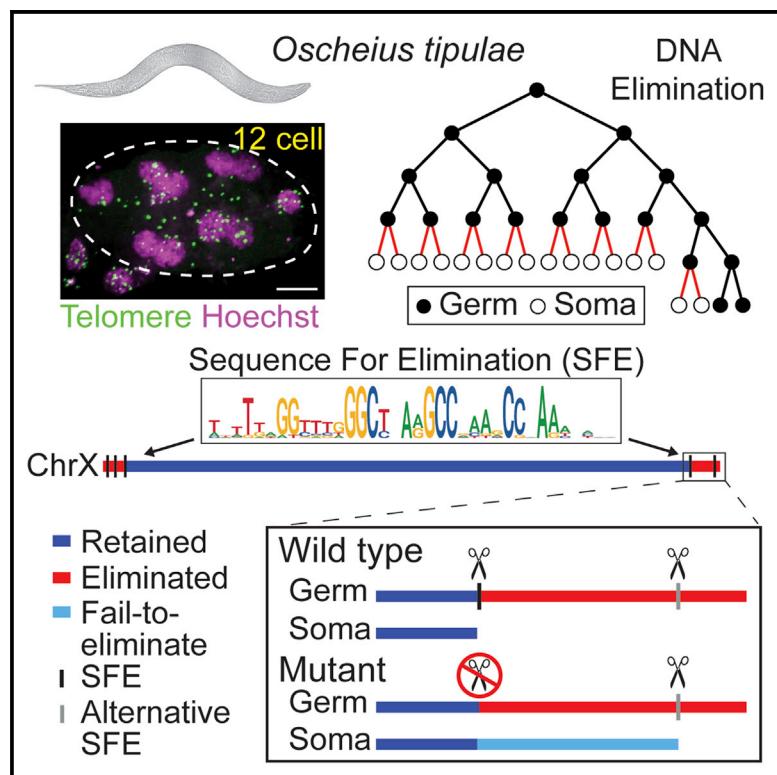


### The nematode *Oscheius tipulae* as a genetic model for programmed DNA elimination

#### Graphical abstract



#### Authors

Thomas C. Dockendorff,  
Brandon Estrem, Jordan Reed, ...,  
Eduardo Villalobos, Erin M. Seaberry,  
Jianbin Wang

#### Correspondence

jianbin.wang@utk.edu

#### In brief

Genetic and molecular analyses in the free-living nematode *Oscheius tipulae* by Dockendorff, Estrem, et al. demonstrate an essential role of a conserved motif for DNA break and a fail-safe mechanism for programmed DNA elimination (PDE). This work establishes *O. tipulae* as a new genetic model for studying mechanisms and consequences of PDE in a metazoan.

#### Highlights

- Programmed DNA elimination (PDE) in *Oscheius tipulae* occurs during 8- to 16-cell stages
- A conserved motif (Sequence For Elimination, SFE) is required for DNA break during PDE
- Additional SFEs act as a fail-safe mechanism to ensure that PDE occurs in *O. tipulae*
- O. tipulae* as a genetic and functional model for PDE in multicellular organisms



## Article

# The nematode *Oscheius tipulae* as a genetic model for programmed DNA elimination

Thomas C. Dockendorff,<sup>1,3</sup> Brandon Estrem,<sup>1,3</sup> Jordan Reed,<sup>1</sup> James R. Simmons,<sup>1</sup> Sobhan Bahrami Zadegan,<sup>2</sup> Maxim V. Zagoskin,<sup>1</sup> Vincent Terta,<sup>1</sup> Eduardo Villalobos,<sup>1</sup> Erin M. Seaberry,<sup>1</sup> and Jianbin Wang<sup>1,2,4,\*</sup>

<sup>1</sup>Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, Knoxville, TN 37996, USA

<sup>2</sup>UT-ORNL Graduate School of Genome Science and Technology, University of Tennessee, Knoxville, TN 37996, USA

<sup>3</sup>These authors contributed equally

<sup>4</sup>Lead contact

\*Correspondence: jianbin.wang@utk.edu

<https://doi.org/10.1016/j.cub.2022.10.043>

## SUMMARY

Programmed DNA elimination (PDE) is a notable exception to the paradigm of genome integrity. In metazoa, PDE often occurs coincident with germline to somatic cell differentiation. During PDE, portions of genomic DNA are lost, resulting in reduced somatic genomes. Prior studies have described the sequences lost, as well as chromosome behavior, during metazoan PDE. However, a system for studying the mechanisms and consequences of PDE in metazoa is lacking. Here, we present a functional and genetic model for PDE in the free-living Rhabditidae nematode *Oscheius tipulae*, a family that also includes *Caenorhabditis elegans*. *O. tipulae* was recently suggested to eliminate DNA. Using staged embryos and DNA FISH, we showed that *O. tipulae* PDE occurs during embryogenesis at the 8–16 cell stages. We identified a conserved motif, named Sequence For Elimination (SFE), for all 12 break sites on the six chromosomes at the junctions of retained and eliminated DNA. SFE mutants exhibited a “fail-to-eliminate” phenotype only at the modified sites. END-seq revealed that breaks can occur at multiple positions within the SFE, with extensive end resection followed by telomere addition to both retained and eliminated ends. We identified many functional SFEs at the chromosome ends through END-seq in the wild-type embryos, genome sequencing of SFE mutants, and comparative genomics of 23 wild isolates. We suggest that these alternative SFEs provide flexibility in the sequences eliminated and a fail-safe mechanism for PDE. These studies establish *O. tipulae* as a new, attractive model for studying the mechanisms and consequences of PDE in a metazoan.

## INTRODUCTION

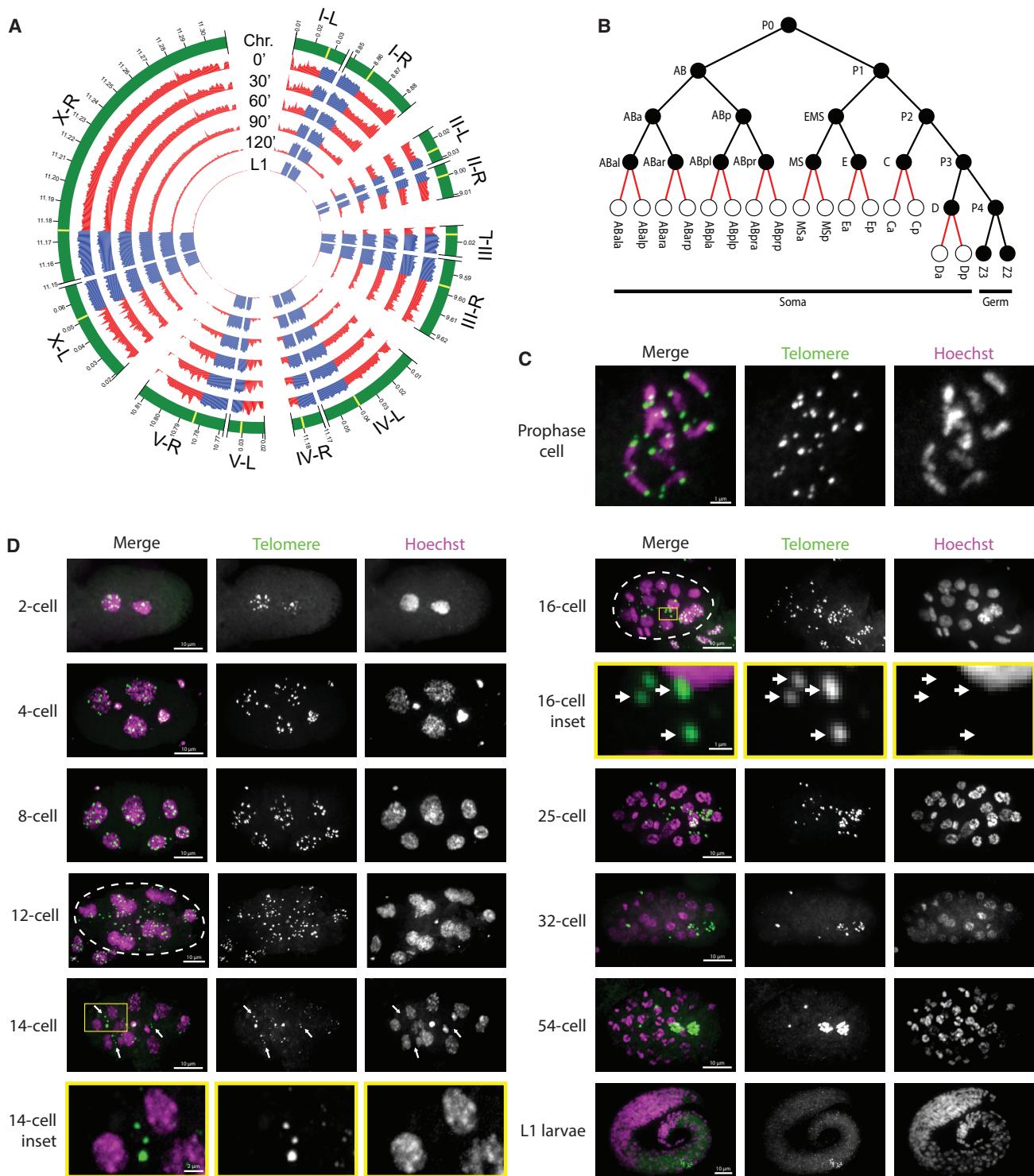
Genome integrity is essential to life. Considerable efforts are made to maintain the stability of genomes, as exemplified by elaborate DNA repair mechanisms, cell cycle checkpoints, faithful segregation of chromosomes during mitosis and meiosis, and suppression of mobile genetic elements.<sup>1–4</sup> Yet genomes also undergo constant changes, often random and small in scale, providing mechanisms for evolution and adaptation. In contrast, programmed DNA elimination (PDE) is a dramatic form of genome change with large amounts of DNA, ranging from 0.5% to 95% of the genome, eliminated during development.<sup>5–9</sup> PDE is highly selective and reproducible and is an integral part of biology for diverse organisms.

PDE was discovered by Boveri<sup>10</sup> in the 1880s, in the horse parasite *Parascaris*. It has since been found in single-cell ciliates,<sup>11,12</sup> in a variety of metazoa across animal phyla,<sup>5,13–16</sup> and in plants.<sup>17,18</sup> Early models for PDE tend to have large genomes and/or discard a large amount of DNA.<sup>5,6</sup> Recently, genomics has been used to characterize metazoan PDE,<sup>19</sup> including the nature of eliminated sequences<sup>20–27</sup> and DNA breaks,<sup>20–22</sup> as well as the evolution of PDE.<sup>15,16,28</sup> A common theme emerging from these studies is the removal of both germline-

expressed genes and repetitive sequences.<sup>5–9</sup> This suggests that one function of PDE in metazoa is to permanently silence certain germline sequences potentially harmful to somatic cells.<sup>6</sup> PDE is also used for sex determination,<sup>7,13,15</sup> remodeling chromosome ends,<sup>22</sup> splitting fused chromosomes,<sup>22</sup> and as a mechanism for evolution.<sup>6,8,9</sup>

Despite the detailed descriptions of PDE in many organisms, functional and mechanistic studies of metazoan PDE are limited.<sup>29,30</sup> Mechanistic insights for PDE, including the contributions and roles of transposable elements and small RNAs, have primarily been gleaned from single-cell ciliates.<sup>11,12,31,32</sup> The features of PDE in metazoa differ considerably from ciliates. Thus, their underlying mechanisms may differ significantly as well.<sup>6–8</sup> Unfortunately, most metazoan models for PDE are not well suited for genetic, functional, and/or biochemical analyses. For example, in one of the best-studied organisms, the human and pig parasite *Ascaris*, genetic manipulation is difficult and propagation of the parasite in the pig host is expensive and not sustainable.<sup>33</sup> Genetic and functional tools available for traditional model organisms are not available or are limited for these PDE models.<sup>7,8</sup> Consequently, concerted genetic, functional, and other molecular analyses have not been carried out on metazoan PDE.



**Figure 1.** *O. tipulae* undergoes PDE during early embryogenesis at 8- to 16-cell stages

(A) Genomic DNA at the ends of *O. tipulae* chromosomes is lost during early development. A Circos<sup>98</sup> plot shows the Illumina read coverage for early embryos (time indicates minutes of development after egg harvest, 0' = 2–4 cells, see Figure S1) and L1 larvae. The retained DNA is blue, and the eliminated DNA is red. For simplicity, the majority of retained DNA is not shown (indicated as a gap, ||, between blue regions). The junctions of retained and eliminated DNA are colored in yellow (outer circle). Genome coverage is plotted in 500-bp bins, and the coordinates are displayed in Mb.

(B) Timing of PDE in the early embryonic blastomere divisions of *O. tipulae*. Filled-in circles represent cells that have not undergone PDE. PDE events are shown using red lines, and cells produced from PDE divisions are represented with empty circles. Cell lineages are named and labeled using the terminology from *C. elegans*.

(legend continued on next page)

Here, we built upon and extended a previous genomic observation<sup>34</sup> and established a genetic and functional model for PDE in the free-living nematode *Oscheius tipulae*, a member of the Rhabditidae family, which also includes *Caenorhabditis elegans*.<sup>35</sup> We showed that PDE occurs during the 8- to 16-cell stages of embryogenesis in *O. tipulae*. Notably, we identified and characterized a conserved sequence motif associated with the DNA break sites and demonstrated its direct role in PDE. DNA breaks occurred within the motif, followed by end resection and telomere healing. Additional breaks occurred simultaneously in the eliminated DNA, perhaps serving as a fail-safe mechanism. We revealed the abundance and variations of this motif in 23 geographically diverse wild isolates of *O. tipulae*. The ability to utilize the available and emerging tools of genomics, biochemistry, molecular biology, cell biology, and genetics, along with its small genome and short life cycle, make *O. tipulae* an excellent model for studying the molecular mechanisms and consequences of PDE in a multicellular organism.

## RESULTS

### *O. tipulae* undergoes PDE during early embryogenesis at 8- to 16-cell stages

*O. tipulae* is best known as a comparative model to *C. elegans* for studies on the evolution of developmental pathways.<sup>35–38</sup> *O. tipulae* was suggested to eliminate DNA in a recent genomic study from The Tree of Life project at the Sanger Institute,<sup>39</sup> where Gonzalez de la Rosa et al. assembled a telomere-to-telomere genome of *O. tipulae* and found a lower coverage of sequencing reads at the ends of all six chromosomes.<sup>34</sup> These data were reminiscent of DNA elimination in the parasitic nematode *Ascaris*, where the ends of all 24 germline chromosomes are removed.<sup>22</sup> The decreased read coverage (compared to the complete loss of reads) was consistent with the analysis of a mixed population of worms (embryos, larvae, and gravid adults) used in the study, thus reflecting the presence of both germline and somatic cells in the samples.<sup>19,20</sup>

To confirm that PDE indeed occurs in *O. tipulae*, we collected (see STAR methods and Figure S1) and sequenced genomes from the germline (mainly 2–4 cell embryos), multiple stages of early embryogenesis, and somatic cells (L1 larvae) (Figure 1). We found that in the 2- to 4-cell sample, the reads at all ends of chromosomes have a higher genome coverage than in later stages of development. The coverage decreases progressively through embryogenesis to ~4% in the L1 stage. These changes in genome coverage are proportional to the percentage of germ cells in the sample. The progressive drop of read coverage

indicates the loss of DNA in the somatic cells, suggesting that PDE occurs during early embryogenesis (Figure 1).

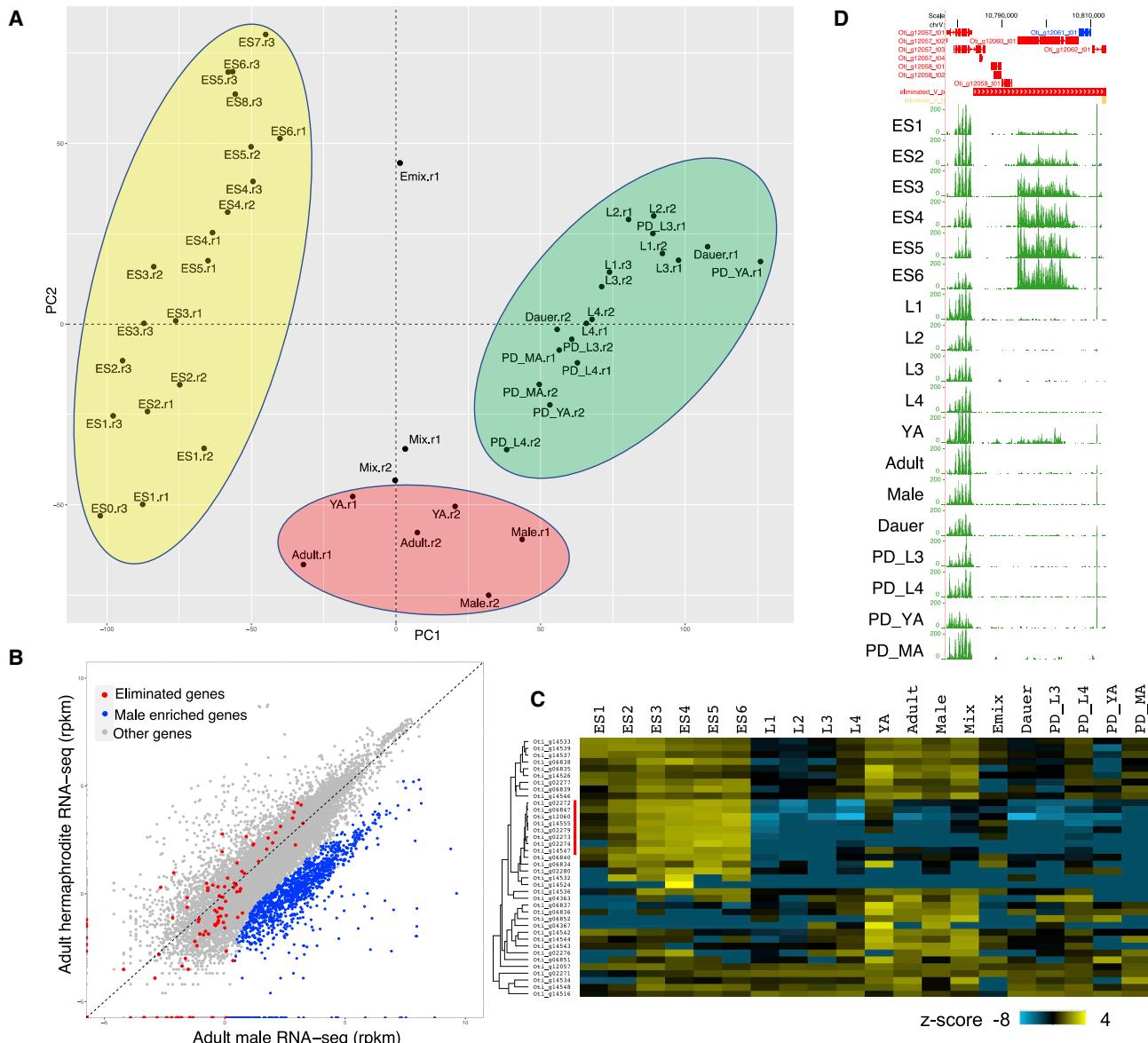
In all studied ascidian nematodes, PDE also occurs during early embryogenesis. However, the exact timing of PDE varies.<sup>40</sup> For example, in *Parascaris*, PDE occurs during the 2- to 32-cell stages, while in *Ascaris*, it occurs in 4- to 16-cell embryos.<sup>13</sup> To determine the precise timing of PDE in *O. tipulae*, we carried out DNA fluorescence *in situ* hybridization (FISH) using a probe against the telomere repeat on early embryos and through L1 larvae (Figures 1C and 1D). Telomere probe signal was visible in all nuclei of embryos up to the 8-cell stage. However, the signal was present in only a few nuclei in 12- or 14-cell embryos and the two presumptive germ cells in a 16-cell embryo (Figure 1D). These data suggest that PDE in *O. tipulae* occurs during the 8- to 16-cell stages (Figure 1B). We also observed that the eliminated DNA remains at the metaphase plate during PDE anaphase (Figure 1D, 14-cell), as observed in *Ascaris*.<sup>22</sup> The eliminated DNA takes one to two cell cycles to degrade (Figure 1D, 25- and 32-cell), and new telomeres in the somatic cells are not detectable by DNA FISH until 32- to 64-cell embryos (Figure 1D, 32- and 54-cell). Interestingly, the lengths for the somatic telomeres were much shorter than the two primordial germ cells (PGCs) as seen in later embryos and L1 larvae (Figure 1D). The telomere FISH probe is much more sensitive than Hoechst dye, as foci of eliminated DNA labeled by the FISH probe were not visible with Hoechst (Figure 1D, 16-cell inset). This suggests that the DNA foci observed with Hoechst staining in *O. tipulae* early embryos (Figure S2) are likely coming from the breakdown of the polar bodies, as demonstrated in *C. elegans*.<sup>41</sup> These sequencing and staining data demonstrate that *O. tipulae* undergoes PDE during embryogenesis at the 8- to 16-cell stages.

### Expression of *O. tipulae* eliminated genes

Previous work identified 115 genes in the eliminated regions, including many helitron-like genes.<sup>34</sup> However, the gene models were primarily based on *in silico* prediction, and expression data were lacking. To determine the expression of retained and eliminated genes, we built comprehensive transcriptomes (RNA sequencing [RNA-seq]) for *O. tipulae* on 22 developmental samples (Figure 2A; Table S2). These samples include early embryogenesis with staged embryos before, during, and after PDE, as well as samples from gastrulation, morphogenesis, staged larvae (L1–L4), dauer, post-dauer larvae, young adults, and mature adults (Figure 2A). Ascarid nematodes eliminate many genes expressed during spermatogenesis.<sup>20–22</sup> Thus, we also sequenced RNA from male worms (handpicked from a high incidence of males [HIM] mutant) to identify male-enriched genes (Tables S3 and S4). We refined

(C) The telomere DNA FISH probe is specific to telomeres. Prophase chromosomes from an early blastomere are shown with telomere probe (green) and Hoechst-33342 DNA stain (magenta). Probe signal is found exclusively at the ends of the chromosomes.

(D) DNA FISH for various developmental timepoints between the 2-cell stage and L1 larvae. Dotted white lines delineate the borders of embryos in images with more than one embryo visible. Yellow boxes show the location of figure insets, which are outlined in yellow. Arrows in the 14-cell embryo show three instances of telomere signals left behind at the metaphase plate between sister nuclei after PDE. The 14-cell inset shows a magnified view of two sister nuclei with eliminated telomere signals in between. The 16-cell inset shows several foci of eliminated DNA in which Hoechst-33342 signal is not visible (white arrows), indicating that Hoechst-33342 staining is not sensitive enough to detect the eliminated DNA. Telomere signals remain in the cytoplasm but are turned over in one to two cell cycles (see 25-cell through 54-cell embryos). Nascent telomeres on newly healed somatic genomes become visible by the 32-cell stage, but they do not grow to the same length as the germline telomeres by the L1 larval stage. See also Figures S1 and S2.



**Figure 2.** *O. tipulae* eliminated DNA encodes germline- and early embryo-expressed genes

(A) Comprehensive RNA-seq for *O. tipulae* embryos and other developmental stages. The principal-component analysis plot shows the relationship of the RNA expression profiles among the sampled developmental stages and replicates (labeled as r1, r2, and r3). Sample descriptions: ES indicates embryo stage, and the numbers of the cells are for the majority of the embryos, with ES0 = 2 cells, ES1 = 2–4 cells, ES2 = 4–8 cells, ES3 = 8–16 cells, ES4 = 16–32 cells, ES5 = 32–50 cells, ES6 = 50–100 cells, ES7 = 100–150 cells, and ES8 = 150–200 cells; PD, post-dauer; YA, young adult; MA, mature adult; Emix, mix of embryos, larvae, and adults (see Figure S1; Table S2). Note the clustering of the embryos (yellow), larvae and post-dauer worms (green), and mature worms (red).

(B) Eliminated genes are not enriched in the males. Comparison of RNA expression (rpkm) between males and hermaphrodites revealed 1,229 male-enriched genes. None of them overlap with eliminated genes (red).

(C) Heatmap showing the expression profiles of 36 eliminated genes. The average rpkm values from the replicates across the stages were converted to Z score. The other 74 eliminated genes have little to no expression across these developmental stages (rpkm < 2). Red vertical line indicates a cluster of eight genes with elevated expression during embryogenesis.

(D) Genome browser view of one of the eight eliminated genes from C. See also Figures S3 and S4 and Tables S2, S3, and S4.

the gene models, extended the 5' and 3' UTRs, identified alternatively spliced isoforms and 560 new genes, and reduced 1,809 genes following the merging/collapsing of gene models (Figure S3). Overall, this RNA-based gene prediction identified 14,558 genes, 36,790 transcripts, and 112 eliminated genes (see Table S3).

One-third (38) of the eliminated genes are expressed in the germline or early embryos (Figure 2C; Table S3). However, the expression level of these genes is, on average, 2.4-fold lower than retained genes (Table S3). This could be due to the reduced level of germline sequences in these mixed samples and/or because most of the eliminated genes are not transcriptionally

active since they are near the heterochromatic telomeres. Consistently, the other two-thirds (74) of the eliminated genes have little or no expression in all sampled stages (maximal RPKM < 2). About half of the eliminated genes are testis specific in ascarids.<sup>20,21,42</sup> Our RNA-seq data from the males identified 1,229 male-enriched genes (5× adult male over mature hermaphrodite, adjusted p value < 0.05; **Figure S4**; **Table S4**), including major sperm proteins and the Argonaute ALG-4.<sup>43,44</sup> However, none of the 112 eliminated genes are enriched in males (**Figure 2B**). Gene ontology (GO) and tissue enrichment analysis<sup>45,46</sup> shows that expression of the eliminated genes is enriched in *C. elegans* Z2 and Z3 germ cells, while no GO enrichment was identified due to the low number of eliminated genes and the presence of many hypothetical genes (**Table S3**).

Surprisingly, eight genes from the eliminated regions are highly expressed in the 64- to 128-cell stage (**Figures 2C** and **2D**). This is a stage where all PDE events have been completed (**Figure 1**). These genes do not have extra copies in the retained regions of the genome. Since these RNAs are not highly expressed in earlier stages, they must be newly transcribed—likely from the two PGCs at a very high level given their small percentage in the sample. Gene annotation indicates that these genes are orthologous to the ATP-dependent DNA helicase (*pif-1*) in *C. elegans* and that they contain a helitron-like domain. Altogether, our RNA-seq data and functional annotations suggest that the eliminated genes are expressed in the germline or early embryos and are not enriched in males. A few genes have a burst of transcription in the PGCs. In addition, many eliminated genes have no predicted functions,<sup>6,47</sup> consistent with observations in other nematode systems,<sup>6,47</sup> suggesting that they may have evolved recently.

#### Differentially expressed genes during *O. tipulae* PDE

We defined differentially expressed genes during early embryogenesis where PDE occurs (**Figure S4**; **Table S4**). We predict that genes involved in PDE will be highly expressed in the 8- to 16-cell embryos but downregulated in the 32- to 128-cell embryos. Using RNA-seq, we identified 622 genes that are consistently enriched (2× fold change and adjusted p value < 0.05) during PDE stages (**Figure S4**; **Table S4**). GO analysis suggests that these genes are enriched in meiotic cell cycle function, macromolecule catabolic processing, reproduction, pole plasm, and ribonucleoprotein granules (**Table S4**). In comparison, 1,317 genes are upregulated in late embryos and are enriched in neuron development, regulation of cell motility, cell morphogenesis, and cell projection organization. Many enriched genes may be involved in early developmental processes. However, we also expect that genes involved in PDE are enriched in the early embryos. For example, the expression of the telomerase gene (*Oti\_g07275*) is upregulated during early embryogenesis and peaks at the 8- to 16-cell stage, consistent with its role in telomere healing during PDE (**Figure 1**). Our comprehensive developmental RNA-seq expression profile provides a critical genomic resource for future genetic, molecular, and developmental studies of PDE and in *O. tipulae* biology in general.

#### A conserved motif at *O. tipulae* break regions

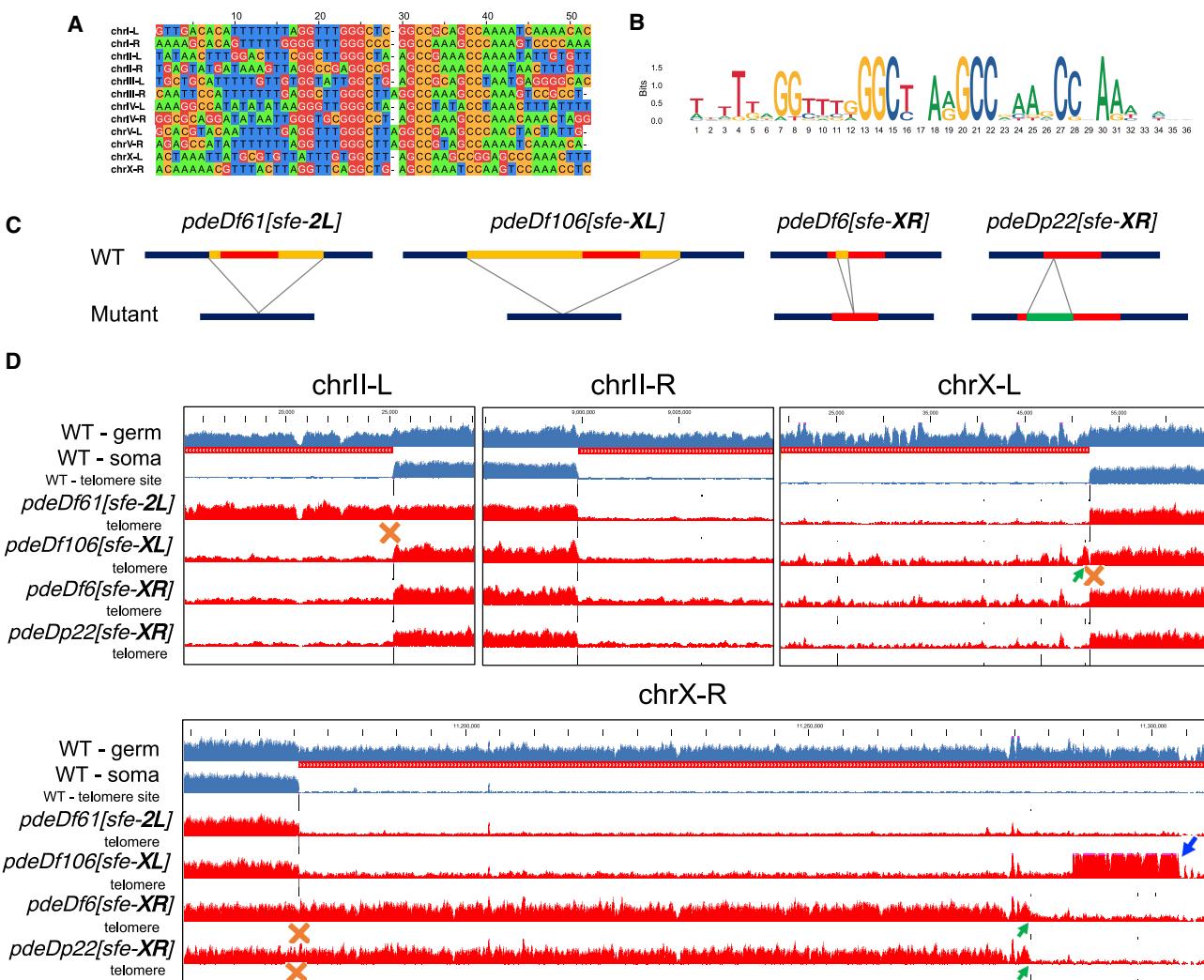
PDE reproducibly leads to DNA breaks at specific regions of nematode chromosomes. How does the cell determine where the DNA breaks should occur? In *Ascaris*, telomere healing, and likely the

DNA break, occurs heterogeneously within 3–6 kb regions called chromosomal breakage regions (CBRs).<sup>20–22</sup> Sequence analysis has identified no conserved motifs within these CBRs.<sup>21</sup> In contrast, examination of *O. tipulae* sequences at the junctions of the retained and eliminated DNA reveals a highly conserved motif (**Figures 3A** and **3B**). This motif, named Sequence For Elimination (SFE, or pronounced as SAFE), spans about 30 bp and is present at the 12 sites on both ends of the six chromosomes. Since these 12 sites determine where the DNA breaks occur and what distal sequences are eliminated in the wild-type CEW1 lab strain, we define these as canonical break sites (see below for alternative sites). The most conserved bases of the motif are arranged in a palindrome (**Figure 3B**)—a common recognition signal for a variety of DNA-binding proteins, including those that cleave or modify DNA. The strong conservation of this motif at the DNA break sites suggests that the SFE motif may participate in the DNA break and/or telomere addition processes.

#### Mutations in the SFE motif result in a fail-to-eliminate phenotype

To explore the role of the SFE motif in the DNA elimination/telomere addition process, we used CRISPR-Cas9 to disrupt the motif at three break sites. These sites were chosen due to their variation in the amount of non-telomeric sequences lost (chrII left = 10.1 kb, chrX left = 32.7 kb, and chrX right = 133.5 kb). Using a “roller” phenotype as a co-CRISPR marker (see **STAR methods** and **Figure S7**), we obtained four SFE mutants with disrupted motifs (**Figure 3C**). These mutants were confirmed by PCR analyses and Sanger sequencing (**Figure S5**). All four SFE homozygous mutants are viable and fertile. However, this does not preclude the possibility of subtle phenotypes in these single mutants or that mutations to additional SFEs might result in more apparent phenotypes. Genome sequencing from mixed populations (eggs, larvae, adults) shows that mutations of this consensus motif block elimination at the altered sites, indicating that the SFE is a key element involved in the elimination process, providing a sequence for the DNA break and telomere addition.

Interestingly, alternative elimination sites within the eliminated DNA were often used in these mutants. These sites, closer to the ends of the chromosomes, led to elimination of smaller regions of the chromosome end (**Figure 3D**). For example, two independent mutations within the SFE motif at the right end of chrX, *pdeDf6[sfe-XR]*, and *pdeDp22[sfe-XR]* resulted in the retention of 106.6 kb of the typically eliminated 133.5 kb of DNA (**Figure 3D**, alternative SFE indicated by green arrows in chrX-R). Furthermore, a mutation that removed the SFE motif from the left end of chrX (*pdeDf106[sfe-XL]*) led to a nearby alternative SFE being used. As a result, only 473 bp escaped PDE (**Figure 3D**, green arrow in chrX-L). The use of these alternative sites suggests that they may serve as a fail-safe mechanism to ensure the removal and remodeling of the chromosome end (see below END-seq results in “Additional break sites in the eliminated regions provide a fail-safe mechanism for PDE”). We note, however, that not all altered SFE sites led to the use of an alternative break site. The SFE mutant on the left end of chrII (*pdeDf61[sfe-2L]*) resulted in the retention of the entire small 10.1 kb subtelomeric region and its telomeric DNA.



**Figure 3. Conserved motif at *O. tipulae* break region is essential for DNA elimination**

(A) Alignment of 12 sequences from the junctions of the retained and eliminated DNA (canonical SFE sites; these sites define the somatic telomere ends in the wild-type CEW1 strain) using Jalview<sup>99</sup> with manual adjustment.

(B) Motif for the canonical SFEs using ggseqlogo.<sup>100</sup>

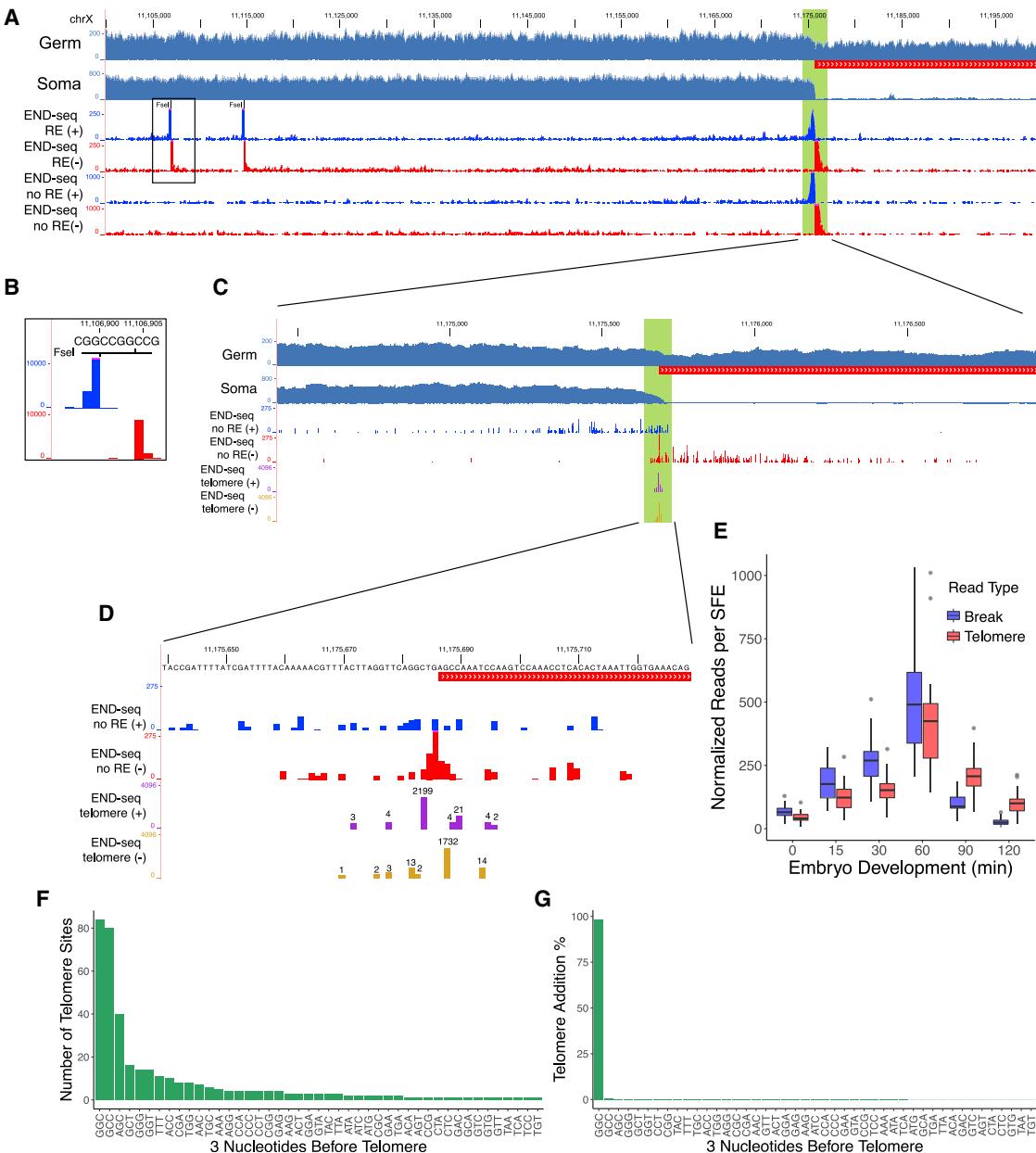
(C) Schematic design of the genome editing for the SFEs. Yellow, deleted DNA; red, core motif region (conserved motif region spanning 30 nt, see Figures 3A and 3B); and green, inserted DNA.

(D) “Fail-to-eliminate” phenotype confirmed by genome sequencing. Four ends of the *O. tipulae* chromosomes are shown (labeled at the top). For each end, genome coverage from the wild-type germline and somatic cells (blue; same data as in Figure 1A, 0' and L1) and sequencing reads from all four mutants (red) are shown. The eliminated regions are marked by red bars below the germline track. Also shown are the telomere addition sites derived from the sequencing data. The mutated sites are indicated with orange Xs. At each mutated site, either a complete (chrII-L) or a partial (chrX-L and chrX-R) “fail-to-eliminate” phenotype was observed. Green arrows point to the alternative sites used in the mutants. The blue arrow indicates DNA with an elevated read coverage that is supposed to be eliminated in the somatic cells. chrII-R serves as a control site where no SFE has been modified. The genome coverage for the eliminated regions varies due to the variations of germ cells percentage in the samples harvested for sequencing. See also Figures S5 and S7 and Table S1.

### *O. tipulae* DNA break ends are resected and healed with telomere addition

To further characterize the DNA breaks in *O. tipulae* PDE, we adapted an END-seq approach that identifies DNA double-strand breaks (DSBs) and the resection of DSB ends.<sup>48,49</sup> We demonstrated that END-seq can identify exogenously introduced FseI or AsI SI DSBs with resected ends (3' overhangs) at single nucleotide resolution in *O. tipulae* early embryos (Figures 4A and 4B). Our END-seq revealed that PDE breaks

occur within the conserved SFE motif, followed by extensive end resections (an average of 1.4 kb) to generate long 3' overhangs (Figure 4C) that are necessary for telomere addition.<sup>50,51</sup> While END-seq captures the short end of the overhang (see STAR methods<sup>48,49</sup>), the long end of the overhang is within the SFEs where breaks occur and telomeres are added (see below). Within the motif, we observed overlapping END-seq signal from forward and reverse strands, suggesting that the DNA breaks can occur at various positions within the motif.



**Figure 4.** *O. tipulae* break ends are resected and healed with telomere added at both ends

(A–D) END-seq data from a population of 4- to 8-cell embryos showing features of the DSBs.

(A) Genome browser view of full END-seq reads from a 100-kb region at the right end of chrX containing two Fsel restriction sites and an SFE site (chrX-R). Germline and somatic read coverage (same data as in Figure 1[A], 0' and L1) are shown to indicate the break site and the eliminated DNA (red line). END-seq libraries with or without Fsel restriction digestion are labeled as RE or no RE. The reads are separated into the forward (+) and reverse strand (-). The END-seq read coverage is shown. Highlighted green region (2.5 kb) at the break site is zoomed in (C).

(B) Inset of one Fsel site. Libraries treated with Fsel show high enrichment of END-seq reads at all Fsel sites. The gap between the two strands is caused by the 4-nt 3' overhang that is removed during END-seq blunting procedure (see STAR methods). The coverage for the end of the reads is shown.

(C) Zoomed-in view of the 5' end of END-seq reads at the SFE. Counts for the end of the reads accumulate in the tails on both sides of the break, consistent with the end resection pattern at a DSB.<sup>48,49</sup> The END-seq signal is more enriched at the core motif region. The overlapping signals from the two strands cannot be derived from a single break site (see Figure 4[B]), indicating that some DSBs can occur at other sites within the overlapping region. Also shown are the telomere addition sites (purple and gold tracks) enriched in the END-seq data (see STAR methods).

(D) Zoomed-in view of a 90-bp region containing the core motif, showing the sequence, 5' ends of END-seq reads and the frequency of telomere-unique reads (in log 2 scale, with the number on the top of the site). Note the close-to-equal frequencies of telomere addition at both the retained and eliminated ends, with the most common telomere addition sites at GGC/GCC.

(E) Delay of telomere addition after the DSBs. The number of reads for END-seq breaks (blue) and telomere addition (red) within 1 kb on either side of the SFE are shown for the early developmental stages. Note that the timing of telomere addition shows a delay relative to DSB formation.

END-seq can also capture new telomere addition events (as long as the telomere length does not exceed the length of Illumina reads; see [STAR methods](#)). Our analysis revealed that telomere addition occurs on both the retained and eliminated ends with equal read frequency ([Figures 4C and 4D](#)), suggesting an initial unbiased telomere healing process to both ends of a break. Strikingly, >98% of the telomere addition sites are within the GGC/GCC position in the center of the palindromic motif ([Figures 4F and 4G](#)). We conclude that the 3' overhang of GGC/GCC sequence serves as the primer for synthesizing new telomeres (GGCTTA/TAAGCC)<sub>n</sub> after the break. The other 2% of the telomere addition sites are within the motif but often have only two nucleotides that match the telomere sequence, presumably providing alternative primer sites for telomerase ([Figures 4F and 4G](#)). The existence of multiple telomere addition sites also explains the overlapping END-seq signals on both strands within the SFEs, suggesting that DNA breaks can occur at multiple sites within the SFE motif ([Figure 4C](#)). In addition, END-seq data through early developmental stages show that the 8- to 16-cell stage has the highest number of breaks ([Figure 4E](#)). These data are consistent with the timing of *O. tipulae* PDE at the 8- to 16-cell stage ([Figure 1](#)). Notably, there is a temporal lag between when the DNA breaks occur and the addition of the telomeres ([Figure 4E](#)), suggesting a potential dynamic telomere length and/or a regulated transition between DNA break and telomere healing.

#### **Additional break sites in the eliminated regions provide a fail-safe mechanism for PDE**

Surprisingly, in addition to the 12 canonical SFE sites at the junctions of retained and eliminated DNA, END-seq also identified 12 alternative sites in the eliminated regions ([Figure 5](#)). These break sites would be largely missed by the genome sequencing of somatic tissues,<sup>19</sup> as they are transient in nature and lost together with the entire eliminated regions in the somatic cells. END-seq enriches these signals by capturing the ends of DNA breaks during PDE. The read frequency, end resection pattern, and features of telomere addition for these 12 alternative SFE sites are the same as the 12 canonical SFE sites ([Figures 5A and 5B](#)). Although the sequences for these alternative SFEs exhibit slight variations, they all share the conserved motif ([Figures 5C and 5D](#)). Interestingly, two groups of alternative SFEs, one with six sites and the other with three, share high sequence similarity within each group, suggesting that they were recently duplicated ([Figures 5B and 5C](#)). The distribution of these alternative sites is biased toward telomeres, although not all chromosome ends have an alternative site ([Figure 5B](#)). Our END-seq data and analyses suggest that during the onset of PDE, all 24 breaks occur simultaneously in the *O. tipulae* genome. These alternative SFEs may act as a fail-safe mechanism for *O. tipulae* PDE to ensure that the chromosome ends are remodeled. Indeed, in our SFE mutants at chrX-R (*pdeDf6[sfe-XR]* and *pdeDp22[sfe-XR]*), one of the alternative SFEs is used when the original SFE

is mutated ([Figures 3D and 5B](#)). Furthermore, in the SFE mutant at chrX-L (*pdeDf106[sfe-XL]*), a site not identified in our END-seq data was used to make an alternative break ([Figures 3D and 5B](#)). In sum, our data suggest that additional DNA break sites exist to provide a fail-safe mechanism for PDE. Future studies are needed to identify factors that determine the choice of SFEs in different scenarios, including various mutants and wild isolates (see below).

#### ***O. tipulae* break sites are not more chromatin accessible during PDE**

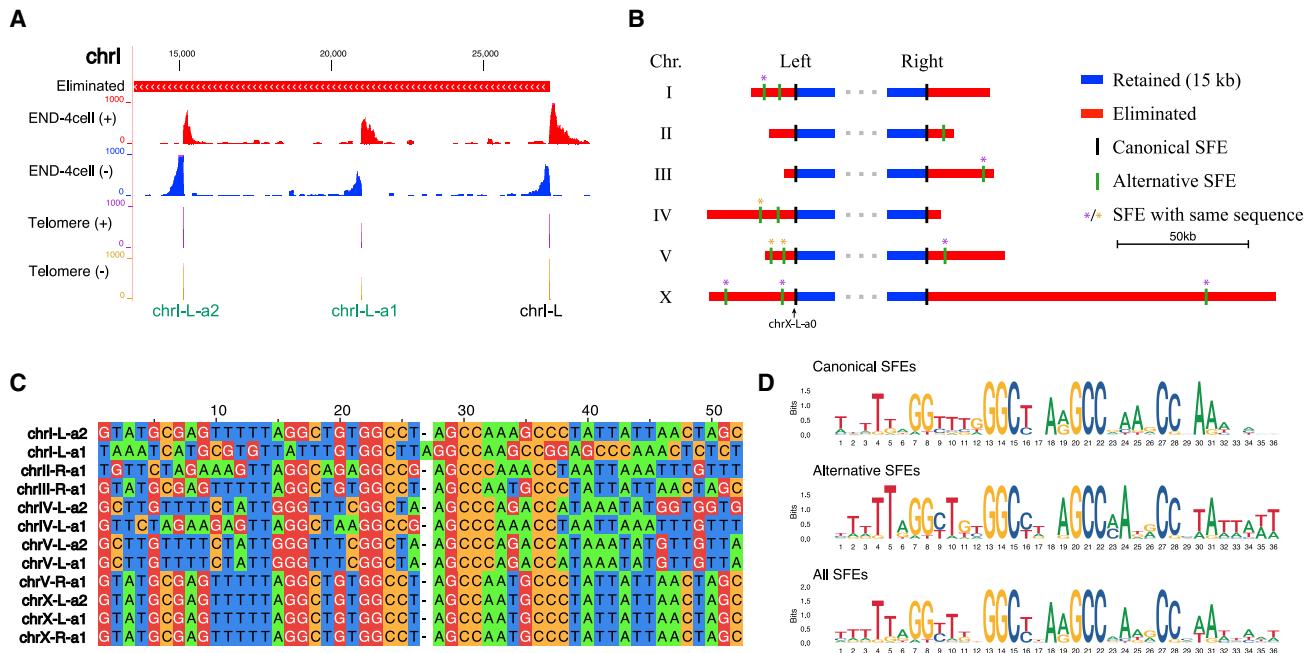
In *Ascaris*, all CBRs become more chromatin accessible during and after PDE.<sup>21,22</sup> We wondered if *O. tipulae* has similar epigenetic features associated with the SFE sites. We used ATAC-seq to determine the chromatin accessibility during *O. tipulae* early embryogenesis ([Figure S6](#)). While ATAC-seq provided a rich trove of information, we did not find an association between open chromatin and the SFEs. The ends of *O. tipulae* chromosomes are, in general, not more chromatin accessible than the middle of the chromosomes. For some sites, the chromatin becomes slightly more accessible after PDE, but this open chromatin is not found across all sites ([Figure S6](#)). This difference in accessibility at the break sites suggests that different mechanisms may be involved for break site recognition and/or DNA breaks in the two species. In *O. tipulae*, the break sites (SFEs) are specifically targeted, while in *Ascaris*, the open chromatin may allow the breaks to occur within a broad region of the CBRs (see [discussion](#)).

#### ***O. tipulae* break sites are flexible in wild isolates**

The breaks and alternative breaks described above are from the CEW1 strain and its mutants. *O. tipulae* has a much broader geographical distribution and a five times higher genetic diversity than *C. elegans*.<sup>52</sup> To assess the conservation and divergence of these break sites in other natural strains, we obtained 23 wild isolates of *O. tipulae* (a gift from Marie-Anne Félix, IBENS; [Table S5](#)) from geographically diverse locations around the world ([Figure 6A](#)). For each isolate, we used Illumina to sequence the genomic DNA from a mixed population (eggs, larvae, and adults) in which ~5%–10% of cells are germline. *De novo* assemblies of the complete mitochondrial genomes were used to build a phylogenetic tree ([Figure 6B](#)). This tree is consistent with the geographic distribution, establishing the evolutionary relationship among the strains.

We mapped the genomic reads from these wild isolates to the CEW1 reference genome to examine the conservation and variation in chromosome end remodeling on the six chromosomes. Two break sites (chrI-L and chrIII-L) show extensive variations among the wild isolates ([Figures 6C and 6D](#)). These variations would result in the elimination of more sequences (1–5 kb) in these wild isolates compared to the CEW1 strain. However, this mapping-based analysis might be biased if the genomes of these strains differ substantially from the reference genome.

(F and G) Sites and frequencies for telomere addition. Plotted are the numbers (F) and frequencies (G) for all *de novo* telomere addition sites captured by END-seq. The DNA for the eliminated side of the motif was reverse complemented to make all 3-mer sequences the same strand as the G-rich strand of the telomere (TTAGGC)<sub>n</sub>. Many sites can be used for telomere addition (F); however, the 3-nt matching GGC is the most frequent site used with 98% of the telomere reads (G). Most of the other 3-mers have two nucleotides that match the telomeric sequence, providing the necessary primer for telomere addition. See also [Figure S6](#) and [Table S1](#).



**Figure 5.** Twelve additional sites in the eliminated regions provide a fail-safe mechanism

(A) END-seq reveals additional break sites in the eliminated regions. Shown is the left end of chrI, where two additional SFEs (chrI-L-a1 and chrI-L-a2) are identified. They have similar END-seq resection patterns and telomere addition reads compared to the canonical SFE (chrI-L).  
 (B) The distribution of the 12 alternative SFEs at the ends of the chromosomes. Most of the retained DNA are not shown (represented by dots). Two groups of SFEs with similar sequences are marked with \* in purple or gold. These are *bona fide* sites as their read frequencies are at the same level (not lower) as the canonical sites. An alternative site, (chrX-L-a0) 473 bp upstream of chrX-L used in the SFE mutant *pdeDf106[sfe-XL]*, is indicated with an arrow. This site was not detected by END-seq.

(C) Alignment of 12 sequences from the alternative SFE sites.

(D) Motifs derived from canonical and/or alternative SFEs. Note the presence of the conserved GGC/GCC sites for all palindromic sequences.

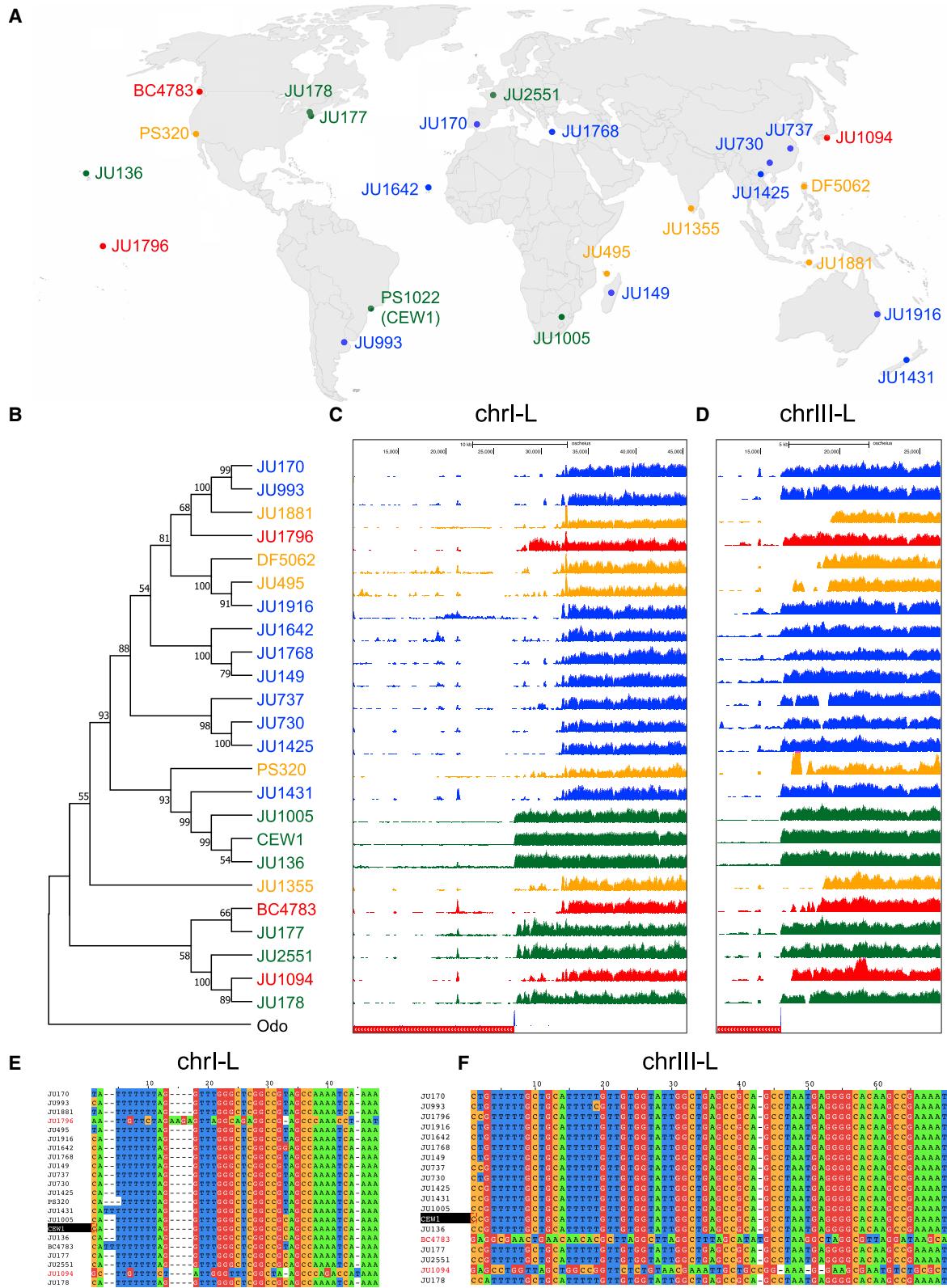
To reduce this potential bias, we used *de novo* assemblies to further evaluate the differences observed (see [STAR methods](#)). We assembled telomere addition sites for each strain at chrl-L and chrlII-L to identify potential SFEs. For the 21 wild isolates with assembled sequences at chrl-L, two (JU1796 and JU1094) have notably different sequences at chrl-L compared to the SFEs from other strains ([Figure 6E](#)). Similarly, two (BC4783 and JU1094) out of the 18 assembled strains for chrlII-L have different SFE sequences ([Figure 6F](#)). While these sequences have deviated from the CEW1 reference at the two sites, they all have the core conserved SFE motif, reinforcing the idea that the motif is essential for the break and telomere addition.

## **Widespread distribution of potential break sites at the ends of CEW1 chromosomes**

We next determined if the divergent SFEs (two for chrl-L and two for chrIII-L, [Figures 6E](#) and [6F](#)) from the wild isolates were present in the CEW1 genome. Surprisingly, all four sequences can be mapped to the CEW1 genome but are in different regions of the CEW1 genome than determined for the wild isolates. This suggests that these sites may have been rearranged (reshuffled) during the evolution of the genomes. This also indicates that the CEW1 genome potentially harbors SFEs that are not used during PDE. Using genomic reads from CEW1 and the other 23 strains, we identified 399 potential SFEs in the CEW1 genome (see [STAR](#)

**methods; Table S6**). Of them, 12 are the canonical breaks, 66 are sites present in multiple strains, 76 are sites unique to a single strain but with high read frequency (>0.5/million reads), and the remaining sites (245) are unique to a single strain and have low read frequency (see examples in Figure 7A). Because these 399 SFEs are from only 24 strains, we predict that the number of potential SFEs will be higher if more strains are examined. Indeed, using FIMO (find individual motif occurrences) predictions,<sup>53</sup> we can identify >10,000 potential sites with the SFE motif across the genome (Table S7), suggesting the possible existence of many more SFEs.

The distribution of these 399 SFEs is highly enriched at the ends of the CEW1 chromosomes ([Figures 7B](#) and [7C](#)). Most of them are in the eliminated regions, close to the canonical break sites, except for chrX-R, where most SFEs are within 30 kb of the end of the chromosome. In addition, the orientation of these SFEs is consistent with the nearby canonical SFEs, and most have the same conserved motif. This suggests a constraint in using the left side of the SFE motif as the retained end. The constraint could be due to their connections to adjacent important retained sequences and/or PDE mechanisms that rely on particular sequences within the motif. The FIMO-predicted sites using the conserved motif showed enrichment at the end of chromosomes, but additional predicted sites were seen throughout the genome, suggesting that sequence alone is not enough to functionally define the sites used for PDE ([Figures 7B](#) and [7C](#)). Collectively, our data illustrate

**Figure 6.** *O. tipulae* break sites are flexible in wild isolates

(A) Wild isolates of *O. tipulae* from around the world used in this study (strains were from M.A. Félix; see Table S5 for descriptions of the strains). The strains were colored based on their phylogenetic relationships and sequence features associated with PDE (see B–F).

(legend continued on next page)

that the *O. tipulae* CEW1 genome is organized to fulfill PDE. The data suggest that many potential SFEs are arranged adjacent to each other in order to ensure that DNA breaks occur at the chromosome ends during PDE.

### DISCUSSION

Programmed DNA elimination was first described by Boveri in the 1880s, yet the molecular details, mechanistic underpinnings, and functional importance of PDE are largely unknown, particularly in multicellular organisms.<sup>5–9</sup> In this study, we illustrated that *O. tipulae*, a free-living nematode with a 60-Mb genome, can be used as a functional and genetic model to study PDE. Through analyses of staged embryos, we determined that PDE occurs during embryogenesis at the 8–16-cell stages. We also analyzed the expression of the eliminated genes and defined features associated with the DNA breaks and telomere healing. A novel discovery for metazoan PDE is the identification of a sequence motif, the SFE, and its direct role in PDE. Our END-seq revealed that in the CEW1 strain, multiple SFEs (12 canonical and 12 alternative) are used simultaneously at the onset of PDE. Additional data from the SFE mutants and wild isolates illustrate that many alternative SFE sites exist and can be used to ensure chromosome end removal and remodeling.<sup>22</sup> The concerted and redundant mechanism for chromosome end remodeling leads us to suggest that PDE likely serves an important function in *O. tipulae*.

#### A comparison of PDE between the nematodes *O. tipulae* and *Ascaris*

Our study allows for a direct comparison of PDE between *O. tipulae* and *Ascaris*, one of the best-studied metazoan models for PDE.<sup>5,6</sup> In both species, PDE occurs during early embryogenesis (4–8- to 16-cell stages). The ends of all chromosomes are broken, resulting in the loss of subtelomeric and telomeric sequences, and new telomeres are added to the broken DNA ends. Recently, PDE has also been demonstrated in other free-living nematodes, including, *Mesorhabditis*,<sup>54</sup> additional *Oscheius* species, and other genera in Rhabditina (P. Gonzalez de la Rosa, L. Stevens, and M. Blaxter, personal communications). Other parasitic nematodes (clade III *Parascaris* and *Toxocara*,<sup>21</sup> *Baylisascaris*,<sup>55</sup> and clade IV *Strongyloides*,<sup>56</sup>) also exhibit DNA breaks and sequence loss, illustrating the broad phylogenetic distribution of PDE in nematodes. However, other well-studied nematodes, including *C. elegans*, do not undergo PDE.<sup>57</sup> Many nematodes have holocentric chromosomes with functional centromeres distributed along the length of the chromosomes,<sup>58</sup> a feature that in principle allows DNA breakage to be more tolerable.<sup>59,60</sup> Holocentric chromosomes also enable centromere reorganization that contributes to the selective segregation of retained DNA during *Ascaris* PDE.<sup>30</sup> These properties may render nematodes more adaptive to PDE.

There are notable differences in PDE between *O. tipulae* and *Ascaris*. *O. tipulae* exhibits no internal chromosome breaks that lead to changes in the number of chromosomes, while many *Ascaris* chromosomes have internal breaks that lead to an increase in its chromosome number after PDE.<sup>22</sup> A second key difference is that *O. tipulae* break (SFE) sites have a distinct motif that is essential for PDE (Figure 3), while *Ascaris* break regions appear to have no common sequence.<sup>21</sup> The most conserved sequence within the *O. tipulae* SFE motif is the GGC/GCC, which likely provides a primer for telomere (TTAGGG) addition. In *Ascaris*, only one nucleotide of homology is needed to prime for *de novo* telomere addition.<sup>21</sup> Thus, the difference in the sequence conservation for the breaks could be in part an adaption to a sequence requirement (or the lack of) for telomere addition. Finally, the CBRs are associated with open chromatin during PDE in *Ascaris*. They remain open after PDE (at 32–256 cells), suggesting that dynamic nucleosome or epigenetic factors are involved.<sup>21</sup> In contrast, *O. tipulae* break regions are not associated with more accessible chromatin and do not show specific accessible changes; they appear to be relatively inaccessible (Figure S6). This suggests that either the chromatin at the break sites is not required to be widely accessible for the break and repair or that the chromatin accessibility is transient in *O. tipulae*. Overall, the differences between *O. tipulae* and *Ascaris* provide a framework to study the variations of PDE mechanisms in both nematodes.

A final difference is that *O. tipulae* eliminates only a very small portion of its genome (0.6% = 349 kb, 112 genes), while *Ascaris* eliminates 18% of its genome (55 Mb, 918 genes). Furthermore, the expression profiles and functional annotation for the eliminated genes have notable differences. *Ascaris* eliminates many genes associated with spermatogenesis,<sup>20–22</sup> while the eliminated genes in *O. tipulae* are not enriched in males. Interestingly, several genes in *O. tipulae* appear to have a burst of expression only in the two PGCs during embryogenesis (Figure 2D). Considering the PGCs are regarded as transcriptionally quiescent in *C. elegans*,<sup>61</sup> it is intriguing to see such a burst of expression for these eliminated genes in *O. tipulae* PGCs. Future studies are required to determine the function of these genes in the PGCs.

#### A fail-safe mechanism for *O. tipulae* PDE

There are 12 canonical SFE sites at the junction of the retained and eliminated DNA on each chromosome that determine the sequences to be eliminated (Figures 1 and 4). Surprisingly, our END-seq revealed 12 additional, alternative SFEs that reside in the eliminated regions (Figure 5). These 24 SFEs are used simultaneously at the onset of PDE. Our mutants with disrupted canonical SFEs confirm that these alternative sites are used (Figure 3D). Although not all eliminated regions contain an alternative site, our data

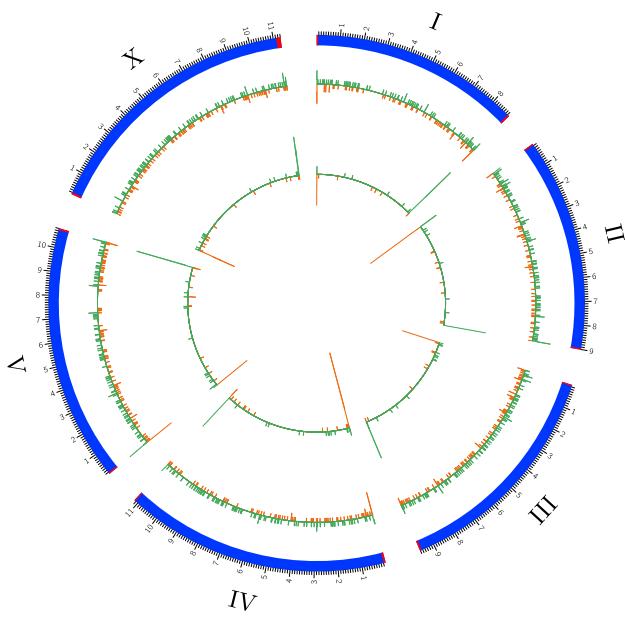
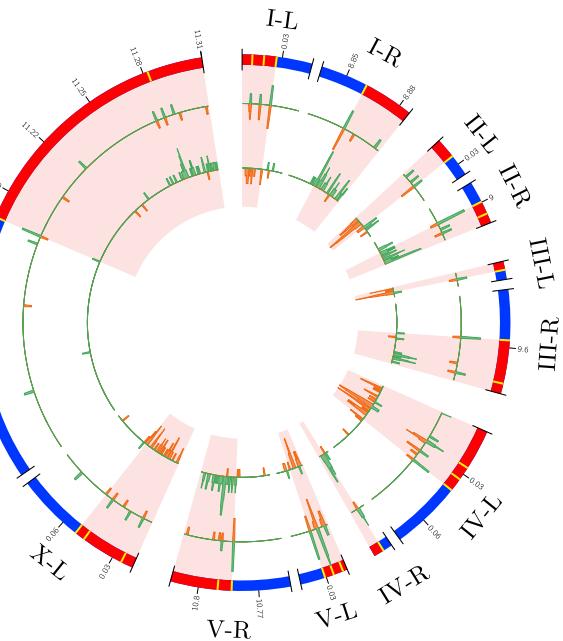
(B) A phylogenetic tree based on the whole mitochondrial genomes. The tree was constructed using MEGA11.<sup>97</sup> The mitochondrial genome of *Oscheius dolichura* (Odo, accession #: OW051498.1) was used as an outgroup for the tree.

(C and D) Variations of the potential break sites and eliminated DNA across the strains. Genomic reads from all strains (including CEW1) were mapped to the CEW1 reference genome. The telomere addition sites (blue) and eliminated DNA (red) for CEW1 are indicated at the bottom. Eliminated DNA in CEW1 genome is marked at the bottom red bar (left sides), with telomere addition sites at the junctions (in blue).

(E and F) Alignment of *de novo* assembled sequences at the break sites for chrl-L (E) and chrIII-L (F). CEW1 is highlighted in black. Strains with notably different sequences to CEW1 are colored in red. Note that not all strains have assembled or identified sequences at these break sites. See also Figure S5.

**A**

SFE Name	END-seq Chr.	Position	JU170	JU893	JU1881	JU1796	DF5062	JU495	JU1916	JU1642	JU1768	JU149	JU37	JU730	JU1425	PS320	JU1431	JU1005	CEW1	JU136	JU1355	BC4783	JU177	JU2651	JU1094	JU178		
chrI-L	Yes	chrI	277185	28	17	10	58	72	69	52	40	10	14	88	28	22	88	81	25	89	100	25	50	2	113	113		
chrII-L	Yes	chrII	886934	35	43	58	72	69	55	40	10	14	88	28	22	88	81	25	89	100	25	50	2	113	113			
chrIII-L	Yes	chrIII	25181	57	89	83	69	96	97	73	80	10	46	56	46	78	85	66	78	81	87	83	64	53	78	79		
chrIV-L	Yes	chrIV	8999759	119	87	127	81	114	131	60	87	36	66	107	71	83	53	83	92	96	89	91	84	70	89	80	114	90
chrIIL-R	Yes	chrIIL	16299	68	62	0	39	0	0	75	68	13	24	81	16	74	0	100	135	106	100	0	0	83	91	0	65	
chrIVL-R	Yes	chrIVL	33	57	78	114	62	75	49	36	43	81	94	65	75	68	63	60	85	106	73	0	2	2	0	65	65	
chrIVR-R	Yes	chrIVR	11179842	34	62	50	71	69	56	55	38	33	50	73	50	65	70	61	54	77	78	53	72	71	80	73	66	
chrIVR-R	Yes	chrIVR	10730448	40	98	87	75	45	92	62	71	30	66	58	45	35	102	102	59	76	60	84	95	83	74	111	73	
chrX-L	Yes	chrX	51049	27	23	30	0	42	27	57	19	23	30	24	20	21	22	29	73	55	24	39	40	23	22	30	31	
chrX-R	Yes	chrX	1175832	69	62	37	33	63	45	58	52	36	46	81	61	48	63	69	65	70	67	80	43	66	40	40		
common_01	chrV	25768	28	33	53	39	30	47	40	33	20	27	29	34	36	22	29	44	33	43	59	43	33	24	45	33		
common_02	chrF	20980	17	21	10	14	9	6	23	9	10	15	32	11	15	4	15	2	2	16	14	16	20	24	8	13		
common_03	chrX	8797764	0	27	0	45	51	0	23	33	69	46	0	0	0	79	0	0	0	0	0	0	0	0	0	0		
common_04	chrF	26257	0	0	0	0	18	0	102	2	7	5	3	0	2	57	0	0	2	1	8	0	0	0	0	0		
common_05	chrF	4999176	80	52	0	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_06	chrIV	9169488	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_07	chrIV	1912371	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_08	chrIV	9377798	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_09	chrIII	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_10	chrX	11096530	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
common_11	chrX	149721	114	37	0	0	0	0	0	23	0	0	0	0	0	0	92	0	0	0	0	0	0	0	0	0		
common_12	chrX	131222	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
Top Unique SFEs	Top Common SFEs																											
unique_01	chrIV	209051	0	0	0	0	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_02	chrF	1011247	0	0	83	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_03	chrF	10511275	0	0	0	0	0	71	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_04	chrF	931026	0	0	0	59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_05	chrF	1316290	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_06	chrF	1316380	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_07	chrF	6177149	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	0	0	0	0	0	
unique_08	chrF	10077565	0	0	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_09	chrF	1832157	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_10	chrF	6438849	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	0	0	0	0	0	0	0	0	0	
unique_11	chrF	7258537	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
unique_12	chrF	10635210	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

**B****C**

**Figure 7. Widespread distribution of potential break sites at the ends of CEW1 chromosomes**

(A) Many new functional SFEs from the wild isolates can be found in CEW1 genome. The CEW1 chromosome coordinates and the normalized telomere-unique reads (out of 10 million reads) from the 24 *O. tipulae* strains are shown in a heatmap. The 12 canonical SFEs, top 12 common SFEs (seen in more than one strains) and top 12 unique SFEs (seen in only one strain) are shown. Some canonical SFEs, including chrI-L, chrII-L, and chrIV-L, do not have telomere reads in several wild strains. Additional SFE sites (399 in total) are in Table S6.

(B and C) Distribution of the SFEs in the CEW1 genome. Circos plot of CEW1 genome with FIMO predicted sites (score > 60; see Table S7 for the full list of sites). These sites are enriched at the chromosome ends but are otherwise randomly distributed across the genome. Shown in the inner circle are the 399 SFEs from the wild isolates that have a highly biased distribution toward the ends of the chromosomes. Green denotes sites on the forward strand and orange shows sites on the reverse strand. Retained DNA is in blue, and eliminated DNA is in red. Genomic coordinates are displayed in Mb. (B) shows the whole chromosome view, while (C) illustrates a zoomed-in view of the eliminated ends (red shaded area; most of the retained DNA is not shown). The canonical and alternative SFEs are indicated in yellow in the outer circle. The majority of the mapped SFEs within the CEW1 eliminated regions (232/241, 96%) have the same strand as used in the wild isolates. See also Tables S6 and S7.

suggest that these alternative sites may serve as a fail-safe mechanism for *O. tipulae* PDE. In addition, genome sequencing and analyses on 23 wild isolates identified ~400

possible alternative SFEs that tend to map to the ends of CEW1 chromosomes (Figure 7). This suggests an evolutionary selection for SFE-like sequences at the ends of the

chromosomes in the *O. tipulae* genome to ensure the removal of the chromosome ends during PDE.

A key question in the field is the function and biological significance of PDE. Our *pdeDf61[sfe-2L]* mutant showed that failure to eliminate a single end of this chromosome led to no visible defects in the worms. However, this mutant only retains 10 kb of unique sequence. In comparison, the *pdeDf6[sfe-XR]* and *pdeDp22[sfe-XR]* mutants failed to eliminate >100 kb of DNA (over 30% of total eliminated DNA in *O. tipulae*) but nevertheless removed the right end of the X chromosome. These mutants also showed no obvious detrimental phenotype, although further careful analyses might reveal subtle phenotypes. It is also possible that retention of multiple chromosome ends could uncover more deleterious phenotypes. We note that the mutants were cultured in a lab environment that may not mimic the challenges that exist in nature. Nevertheless, future work on mutants that fail to eliminate the full end of other single sites, multiple sites, or all ends of chromosomes promises to reveal novel insights into the function of PDE in *O. tipulae*.

Why does CEW1 only make breaks at these 24 sites out of the ~400 potential SFEs? How are these sites selected? Once chosen, how are the breaks made and processed before telomere addition? We hypothesize that the conserved motif sequence (GGC/GCC) may play an essential role during the PDE process. Our SFE mutants showed that disruption of these sequences leads to a fail-to-eliminate phenotype (Figure 3). The SFE is a degenerate palindromic sequence that could be recognized by a DNA-binding protein (Figures 3 and 5). Once bound, the DNA-binding protein can generate the DNA breaks directly or by recruiting other proteins to the sites. In these scenarios, additional factors would be necessary to distinguish these key 24 sites from other potential non-functional sites. These factors could be related to the binding affinity of the DNA/protein interactions. Additional factors might include the 3D genome organization and/or other epigenetic features. It is also possible that the recognition of the break sites is mediated by *cis*- and *trans*-regulatory RNAs (small RNAs or lncRNAs) or other epigenetic factors without the involvement of a specific DNA-binding protein.<sup>62</sup> Alternatively, DNA replication stress, RNA transcription, and R-loops<sup>63–65</sup> induced by the SFEs may contribute to the generation of the DNA breaks. Surprisingly, the chrX-L mutant (*pdeDf106[sfe-XL]*) used a novel site nearby (473 bp), instead of the two existing alternative SFE sites, in response to a disrupted canonical SFE site (Figures 3D and 5B). This indicates that this novel site has the potential to be a break site, but in the wild type, the use of the adjacent canonical break prevents it from serving as an SFE.

### Conclusions

PDE contradicts the genome constancy rule, yet it is seen in many phylogenetically divergent groups, including multiple metazoan, suggesting its broad molecular and biological significance. The molecular mechanisms of PDE in multicellular organisms remain largely unknown, partly due to the challenges of working with organisms that are not traditional models and that have limited tools. Here, we present a new model to study PDE in the free-living nematode *O. tipulae*. Our genomic and functional data reveal the presence of a conserved motif (SFE) that directly facilitates PDE, a feature that has yet to be described for other metazoan PDE models. Our genetic and molecular analyses

show that alternative SFEs can be utilized when necessary, perhaps serving as a fail-safe mechanism for PDE. The amenability to genetic and functional manipulations, short life cycle, modest amount of DNA eliminated, and the ability to capture discrete embryonic stages, including those stages that undergo PDE, makes *O. tipulae* an excellent model to carry out in-depth molecular studies that will uncover the functions, mechanisms, and consequences of PDE in a multicellular organism.

### 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
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Worm strains, culture and maintenance
- **METHOD DETAILS**
  - Egg isolation, staged embryos and larvae, and other samples
  - Hoechst staining
  - DNA FISH
  - RNA isolation and sequencing
  - Genomic DNA extraction and sequencing
  - CRISPR-Cas9 modification of the *O. tipulae* genome
  - Mutant strains nomenclature
  - Identification of a *rol-6* orthologue in *O. tipulae*
  - END-seq library preparation
  - ATAC-seq library preparation
  - Bioinformatic data analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cub.2022.10.043>.

### ACKNOWLEDGMENTS

We thank CGC for *O. tipulae* strains; Aurélien Richaud and Marie-Anne Félix for the wild isolates and sharing unpublished sequencing data; Pablo Manuel Gonzalez de la Rosa and Mark Blaxter for personal communications and sharing genomic data; Sally Adams and Andre Pires da Silva for liposome-based CRISPR methods; and Chris Turpin, Jenny Heppert, Guy Caldwell, Laura Berkowitz, and Tom Evans for worm methods, protocols, and suggestions. We also thank graduate students Jansirani Srinivasan and Yingjie Xu and undergraduate students Ryan Qiu, Mollie Sterling, Gracie Chiampas, Matthew Carr, and Jordan Parker for helping with worm maintenance, sample collection, and/or genetic analysis; the University of Colorado Anschutz Medical Campus Genomics Core for sequencing services; and Mariano Labrador, Albrecht von Arnim, and Dick Davis for comments and critical reading of the manuscript. This work was supported by NIH grant AI155588 and the University of Tennessee Knoxville Startup Funds.

### AUTHOR CONTRIBUTIONS

T.C.D. and J.W. designed the project; J.R. and J.W. developed protocols and prepared staged embryos; J.R.S. and E.V. did staining, DNA FISH, and

imaging; J.R. carried out RNA-seq; T.C.D. and V.T. performed CRISPR and genetic analyses; T.C.D. and E.S. sequenced genomes from staged embryos, mutants, and wild isolates; B.E. did END-seq; J.R.S. did ATAC-seq; B.E., S.B.Z., M.V.Z., and J.W. carried out bioinformatic analyses; and T.C.D., B.E., and J.W. wrote the manuscript.

#### DECLARATION OF INTERESTS

The authors declare no competing interests.

#### INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in their field of research or within their geographical location. We support inclusive, diverse, and equitable conduct of research.

Received: August 22, 2022

Revised: October 5, 2022

Accepted: October 19, 2022

Published: November 14, 2022

#### REFERENCES

- Kolodner, R.D., Putnam, C.D., and Myung, K. (2002). Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* 297, 552–557. <https://doi.org/10.1126/science.1075277>.
- Sancar, A., Lindsey-Boltz, L.A., Unsöl-Kaçmaz, K., and Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annu. Rev. Biochem.* 73, 39–85. <https://doi.org/10.1146/annurev.biochem.73.011303.073723>.
- Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu. Rev. Genet.* 35, 673–745. <https://doi.org/10.1146/annurev.genet.35.102401.091334>.
- Fedoroff, N.V. (2012). Transposable elements, epigenetics, and genome evolution. *Science* 338, 758–767. <https://doi.org/10.1126/science.338.6108.758>.
- Wang, J., and Davis, R.E. (2014). Programmed DNA elimination in multicellular organisms. *Curr. Opin. Genet. Dev.* 27, 26–34. <https://doi.org/10.1016/j.gde.2014.03.012>.
- Zagoskin, M.V., and Wang, J. (2021). Programmed DNA elimination: silencing genes and repetitive sequences in somatic cells. *Biochem. Soc. Trans.* 49, 1891–1903. <https://doi.org/10.1042/BST20190951>.
- Dedukh, D., and Krasikova, A. (2021). Delete and survive: strategies of programmed genetic material elimination in eukaryotes. *Biol. Rev. Camb. Phil. Soc.* 97, 195–216. <https://doi.org/10.1111/brv.12796>.
- Drotos, K.H.I., Zagoskin, M.V., Kess, T., Gregory, T.R., and Wyngaard, G.A. (2022). Throwing away DNA: programmed downsizing in somatic nuclei. *Trends Genet.* 38, 483–500. <https://doi.org/10.1016/j.tig.2022.02.003>.
- Kloc, M., Kubiak, J.Z., and Ghobrial, R.M. (2022). Natural genetic engineering: A programmed chromosome/DNA elimination. *Dev. Biol.* 486, 15–25. <https://doi.org/10.1016/j.ydbio.2022.03.008>.
- Boveri, T. (1887). Ueber Differenzierung der Zellkerne während der Furchung des Eies von *Ascaris megalocephala*. *Anat. Anzeiger* 2, 688–693.
- Chalker, D.L., and Yao, M.C. (2011). DNA elimination in ciliates: transposon domestication and genome surveillance. *Annu. Rev. Genet.* 45, 227–246. <https://doi.org/10.1146/annurev-genet-110410-132432>.
- Bracht, J.R., Fang, W., Goldman, A.D., Dolzenko, E., Stein, E.M., and Landweber, L.F. (2013). Genomes on the edge: programmed genome instability in ciliates. *Cell* 152, 406–416. <https://doi.org/10.1016/j.cell.2013.01.005>.
- Streit, A., Wang, J., Kang, Y., and Davis, R.E. (2016). Gene silencing and sex determination by programmed DNA elimination in parasitic nematodes. *Curr. Opin. Microbiol.* 32, 120–127. <https://doi.org/10.1016/j.mib.2016.05.012>.
- Smith, J.J., Timoshevskiy, V.A., and Saraceno, C. (2021). Programmed DNA elimination in vertebrates. *Annu. Rev. Anim. Biosci.* 9, 173–201. <https://doi.org/10.1146/annurev-animal-061220-023220>.
- Hodson, C.N., and Ross, L. (2021). Evolutionary perspectives on germline-restricted chromosomes in flies (Diptera). *Genome Biol. Evol.* 13, evab072. <https://doi.org/10.1093/gbe/evab072>.
- Borodin, P., Chen, A., Forstmeier, W., Fouché, S., Malinovskaya, L., Pei, Y., Reifová, R., Ruiz-Ruano, F.J., Schlebusch, S.A., Sotelo-Muñoz, M., et al. (2022). Mendelian nightmares: the germline-restricted chromosome of songbirds. *Chromosome Res.* 30, 255–272. <https://doi.org/10.1007/s10577-022-09688-3>.
- Ruban, A., Schmutzler, T., Wu, D.D., Fuchs, J., Boudichevskaia, A., Rubtsova, M., Pistrick, K., Melzer, M., Himmelbach, A., Schubert, V., et al. (2020). Supernumerary B chromosomes of *Aegilops speltoides* undergo precise elimination in roots early in embryo development. *Nat. Commun.* 11, 2764. <https://doi.org/10.1038/s41467-020-16594-x>.
- Blavet, N., Yang, H., Su, H., Solanský, P., Douglas, R.N., Karafiátová, M., Šimková, L., Zhang, J., Liu, Y., Hou, J., et al. (2021). Sequence of the supernumerary B chromosome of maize provides insight into its drive mechanism and evolution. *Proc. Natl. Acad. Sci. USA* 118, e2104254118. <https://doi.org/10.1073/pnas.2104254118>.
- Wang, J. (2021). Genome Analysis of Programmed DNA Elimination in Parasitic Nematodes. *Methods Mol. Biol.* 2369, 251–261. [https://doi.org/10.1007/978-1-0716-1681-9\\_14](https://doi.org/10.1007/978-1-0716-1681-9_14).
- Wang, J., Mitreva, M., Berriman, M., Thorne, A., Magrini, V., Koutsopoulos, G., Kumar, S., Blaxter, M.L., and Davis, R.E. (2012). Silencing of germline-expressed genes by DNA elimination in somatic cells. *Dev. Cell* 23, 1072–1080. <https://doi.org/10.1016/j.devcel.2012.09.020>.
- Wang, J., Gao, S., Mostovoy, Y., Kang, Y., Zagoskin, M., Sun, Y., Zhang, B., White, L.K., Easton, A., Nutman, T.B., et al. (2017). Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* 27, 2001–2014. <https://doi.org/10.1101/gr.225730.117>.
- Wang, J., Veronezi, G.M.B., Kang, Y., Zagoskin, M., O'Toole, E.T., and Davis, R.E. (2020). Comprehensive Chromosome End Remodeling during Programmed DNA Elimination. *Curr. Biol.* 30, 3397–3413.e4. <https://doi.org/10.1016/j.cub.2020.06.058>.
- Sun, C., Wyngaard, G., Walton, D.B., Wichman, H.A., and Mueller, R.L. (2014). Billions of basepairs of recently expanded, repetitive sequences are eliminated from the somatic genome during copepod development. *BMC Genom.* 15, 186. <https://doi.org/10.1186/1471-2164-15-186>.
- Smith, J.J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M.C., Parker, H.J., Cook, M.E., Hess, J.E., Narum, S.R., Lamanna, F., et al. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* 50, 270–277. <https://doi.org/10.1038/s41588-017-0036-1>.
- Biederman, M.K., Nelson, M.M., Asalone, K.C., Pedersen, A.L., Saldaña, C.J., and Bracht, J.R. (2018). Discovery of the First Germline-Restricted Gene by Subtractive Transcriptomic Analysis in the Zebra Finch, *Taeniopygia guttata*. *Curr. Biol.* 28, 1620–1627.e5. <https://doi.org/10.1016/j.cub.2018.03.067>.
- Kinsella, C.M., Ruiz-Ruano, F.J., Dion-Côté, A.M., Charles, A.J., Gossmann, T.I., Cabrero, J., Kappei, D., Hemmings, N., Simons, M.J.P., Camacho, J.P.M., et al. (2019). Programmed DNA elimination of germline development genes in songbirds. *Nat. Commun.* 10, 5468. <https://doi.org/10.1038/s41467-019-13427-4>.
- Hodson, C.N., Jaron, K.S., Gerbi, S., and Ross, L. (2022). Gene-rich germline-restricted chromosomes in black-winged fungus gnats evolved through hybridization. *PLoS Biol.* 20, e3001559. <https://doi.org/10.1371/journal.pbio.3001559>.
- Torgasheva, A.A., Malinovskaya, L.P., Zadesenets, K.S., Karamysheva, T.V., Kizilova, E.A., Akberdina, E.A., Pristyazhnyuk, I.E., Shnaider, E.P., Volodkina, V.A., Saifitdinova, A.F., et al. (2019). Germline-restricted

- chromosome (GRC) is widespread among songbirds. *Proc. Natl. Acad. Sci. USA* 116, 11845–11850. <https://doi.org/10.1073/pnas.1817373116>.
29. Seidl, C., and Moritz, K.B. (1998). A novel UV-damaged DNA binding protein emerges during the chromatin-eliminating cleavage period in *Ascaris suum*. *Nucleic Acids Res.* 26, 768–777. <https://doi.org/10.1093/nar/26.3.768>.
30. Kang, Y., Wang, J., Neff, A., Kratzer, S., Kimura, H., and Davis, R.E. (2016). Differential Chromosomal Localization of Centromeric Histone CENP-A Contributes to Nematode Programmed DNA Elimination. *Cell Rep.* 16, 2308–2316. <https://doi.org/10.1016/j.celrep.2016.07.079>.
31. Noto, T., and Mochizuki, K. (2017). Whats, hows and whys of programmed DNA elimination in Tetrahymena. *Open Biol.* 7, 170172. <https://doi.org/10.1098/rsob.170172>.
32. Rzeszutek, I., Maurer-Alcalá, X.X., and Nowacki, M. (2020). Programmed genome rearrangements in ciliates. *Cell. Mol. Life Sci.* 77, 4615–4629. <https://doi.org/10.1007/s00018-020-03555-2>.
33. Wang, J., and Davis, R.E. (2020). Ascaris. *Curr. Biol.* 30, R423–R425. <https://doi.org/10.1016/j.cub.2020.02.064>.
34. Gonzalez de la Rosa, P.M., Thomson, M., Trivedi, U., Tracey, A., Tandonnet, S., and Blaxter, M. (2021). A telomere-to-telomere assembly of *Oscheius tipulæ* and the evolution of rhabditid nematode chromosomes. *G3 (Bethesda)* 11, jkaa020. <https://doi.org/10.1093/g3journal/jkaa020>.
35. Felix, M.A. (2006). *Oscheius tipulæ*. Worm, 1–8. <https://doi.org/10.1895/wormbook.1.119.1>.
36. Sommer, R.J., and Sternberg, P.W. (1995). Evolution of cell lineage and pattern formation in the vulval equivalence group of rhabditid nematodes. *Dev. Biol.* 167, 61–74. <https://doi.org/10.1006/dbio.1995.1007>.
37. Delattre, M., and Félix, M.A. (2001). Polymorphism and evolution of vulval precursor cell lineages within two nematode genera, *Caenorhabditis* and *Oscheius*. *Curr. Biol.* 11, 631–643. [https://doi.org/10.1016/s0960-9822\(01\)00202-0](https://doi.org/10.1016/s0960-9822(01)00202-0).
38. Besnard, F., Koutsovoulos, G., Dieudonné, S., Blaxter, M., and Félix, M.A. (2017). Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode *Oscheius tipulæ* To Identify Mutations Affecting Vulva Development. *Genetics* 206, 1747–1761. <https://doi.org/10.1534/genetics.117.203521>.
39. Blaxter, M., Archibald, J.M., Childers, A.K., Coddington, J.A., Crandall, K.A., Di Palma, F., Durbin, R., Edwards, S.V., Graves, J.A.M., Hackett, K.J., et al. (2022). Why sequence all eukaryotes? *Proc. Natl. Acad. Sci. USA* 119, e2115636118. <https://doi.org/10.1073/pnas.2115636118>.
40. Müller, F., and Tobler, H. (2000). Chromatin diminution in the parasitic nematodes *ascaris suum* and *parascaris univalens*. *Int. J. Parasitol.* 30, 391–399. [https://doi.org/10.1016/s0020-7519\(99\)00199-x](https://doi.org/10.1016/s0020-7519(99)00199-x).
41. Fazeli, G., Stetter, M., Lisack, J.N., and Wehman, A.M. (2018). *C. elegans* Blastomeres Clear the Corpse of the Second Polar Body by LC3-Associated Phagocytosis. *Cell Rep.* 23, 2070–2082. <https://doi.org/10.1016/j.celrep.2018.04.043>.
42. Wang, J. (2021). Genomics of the Parasitic Nematode *Ascaris* and Its Relatives. *Genes* 12, 493. <https://doi.org/10.3390/genes12040493>.
43. Wang, J., Czech, B., Crunk, A., Wallace, A., Mitreva, M., Hannon, G.J., and Davis, R.E. (2011). Deep small RNA sequencing from the nematode *Ascaris* reveals conservation, functional diversification, and novel developmental profiles. *Genome Res.* 21, 1462–1477. <https://doi.org/10.1101/gr.121426.111>.
44. Zagorskin, M.V., Wang, J., Neff, A.T., Veronezi, G.M.B., and Davis, R.E. (2022). Small RNA pathways in the nematode *Ascaris* in the absence of piRNAs. *Nat. Commun.* 13, 837. <https://doi.org/10.1038/s41467-022-28482-7>.
45. Angeles-Albores, D., N Lee, R.Y., Chan, J., and Sternberg, P.W. (2016). Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinf.* 17, 366. <https://doi.org/10.1186/s12859-016-1229-9>.
46. Angeles-Albores, D., Lee, R., Chan, J., and Sternberg, P. (2018). Two new functions in the WormBase Enrichment Suite. *MicroPubl. Biol.* 2018, <https://doi.org/10.17912/W25Q2N>.
47. Rödelsperger, C., Ebbing, A., Sharma, D.R., Okumura, M., Sommer, R.J., and Korswagen, H.C. (2021). Spatial Transcriptomics of Nematodes Identifies Sperm Cells as a Source of Genomic Novelty and Rapid Evolution. *Mol. Biol. Evol.* 38, 229–243. <https://doi.org/10.1093/molbev/msaa207>.
48. Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B.P., and Nussenzweig, A. (2016). DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol. Cell* 63, 898–911. <https://doi.org/10.1016/j.molcel.2016.06.034>.
49. Wong, N., John, S., Nussenzweig, A., and Canela, A. (2021). END-seq: An Unbiased, High-Resolution, and Genome-Wide Approach to Map DNA Double-Strand Breaks and Resection in Human Cells. *Methods Mol. Biol.* 2153, 9–31. [https://doi.org/10.1007/978-1-0716-0644-5\\_2](https://doi.org/10.1007/978-1-0716-0644-5_2).
50. Lingner, J., and Cech, T.R. (1996). Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang. *Proc. Natl. Acad. Sci. USA* 93, 10712–10717. <https://doi.org/10.1073/pnas.93.20.10712>.
51. Henderson, E.R., and Blackburn, E.H. (1989). An overhanging 3' terminus is a conserved feature of telomeres. *Mol. Cell Biol.* 9, 345–348. <https://doi.org/10.1128/mcb.9.1.345-348.1989>.
52. Baïle, D., Barrière, A., and Félix, M.A. (2008). *Oscheius tipulæ*, a widespread hermaphroditic soil nematode, displays a higher genetic diversity and geographical structure than *Caenorhabditis elegans*. *Mol. Ecol.* 17, 1523–1534. <https://doi.org/10.1111/j.1365-294X.2008.03697.x>.
53. Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. <https://doi.org/10.1093/bioinformatics/btr064>.
54. Rey, C., Launay, C., Wenger, E., and Delattre, M. (2022). Programmed-DNA Elimination in the free-living nematodes *Mesorhabditis*. Preprint at bioRxiv. <https://doi.org/10.1101/2022.03.19.484980>.
55. Xie, Y., Wang, S., Wu, S., Gao, S., Meng, Q., Wang, C., Lan, J., Luo, L., Zhou, X., Xu, J., et al. (2021). Genome of the Giant Panda Roundworm Illuminates Its Host Shift and Parasitic Adaptation. *Dev. Reprod. Biol.* <https://doi.org/10.1016/j.gpb.2021.08.002>.
56. Nemetschke, L., Eberhardt, A.G., Hertzberg, H., and Streit, A. (2010). Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus *Strongyloides*. *Curr. Biol.* 20, 1687–1696. <https://doi.org/10.1016/j.cub.2010.08.014>.
57. Emmons, S.W., Klass, M.R., and Hirsh, D. (1979). Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 76, 1333–1337. <https://doi.org/10.1073/pnas.76.3.1333>.
58. Carlton, P.M., Davis, R.E., and Ahmed, S. (2022). Nematode chromosomes. *Genetics* 221, <https://doi.org/10.1093/genetics/iyc014>.
59. Maddox, P.S., Oegema, K., Desai, A., and Cheeseman, I.M. (2004). Holo"er than thou: chromosome segregation and kinetochore function in *C. elegans*. *Chromosome Res.* 12, 641–653. <https://doi.org/10.1023/B:CHRO.0000036588.42225.2f>.
60. Melters, D.P., Palilis, L.V., Korf, I.F., and Chan, S.W.L. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. *Chromosome Res.* 20, 579–593. <https://doi.org/10.1007/s10577-012-9292-1>.
61. Seydoux, G., and Dunn, M.A. (1997). Transcriptionally repressed germ cells lack a subpopulation of phosphorylated RNA polymerase II in early embryos of *Caenorhabditis elegans* and *Drosophila melanogaster*. *Development* 124, 2191–2201. <https://doi.org/10.1242/dev.124.11.2191>.
62. Betermier, M., and Duharcourt, S. (2014). Programmed Rearrangement in Ciliates: *Paramecium*. *Microbiol. Spectr.* 2, <https://doi.org/10.1128/microbiolspec.MDNA3-0035-2014>.
63. Aguilera, A., and García-Muse, T. (2012). R loops: from transcription by-products to threats to genome stability. *Mol. Cell* 46, 115–124. <https://doi.org/10.1016/j.molcel.2012.04.009>.
64. Zeman, M.K., and Cimprich, K.A. (2014). Causes and consequences of replication stress. *Nat. Cell Biol.* 16, 2–9. <https://doi.org/10.1038/ncb2897>.

65. Petermann, E., Lan, L., and Zou, L. (2022). Sources, resolution and physiological relevance of R-loops and RNA-DNA hybrids. *Nat. Rev. Mol. Cell Biol.* 23, 521–540. <https://doi.org/10.1038/s41580-022-00474-x>.
66. Stiernagle, T. (2006). Maintenance of *C. elegans*. WormBook, 1–11. <https://doi.org/10.1895/wormbook.1.101.1>.
67. Hibshman, J.D., Webster, A.K., and Baugh, L.R. (2021). Liquid-culture protocols for synchronous starvation, growth, dauer formation, and dietary restriction of *Caenorhabditis elegans*. *STAR Protoc.* 2, 100276. <https://doi.org/10.1101/j.xpro.2020.100276>.
68. Wang, J., Garrey, J., and Davis, R.E. (2014). Transcription in pronuclei and one- to four-cell embryos drives early development in a nematode. *Curr. Biol.* 24, 124–133. <https://doi.org/10.1016/j.cub.2013.11.045>.
69. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. <https://doi.org/10.1038/nmeth.2019>.
70. Portman, D. (2006). Profiling *C. elegans* gene expression with DNA microarrays. WormBook, 1–11. <https://doi.org/10.1895/wormbook.1.104.1>.
71. Paix, A., Folkmann, A., and Seydoux, G. (2017). Precision genome editing using CRISPR-Cas9 and linear repair templates in *C. elegans*. *Methods* 121–122, 86–93. <https://doi.org/10.1016/j.jymeth.2017.03.023>.
72. Ghanta, K.S., Ishidate, T., and Mello, C.C. (2021). Microinjection for precision genome editing in *Caenorhabditis elegans*. *STAR Protoc.* 2, 100748. <https://doi.org/10.1101/j.xpro.2021.100748>.
73. Adams, S., Pathak, P., Shao, H., Lok, J.B., and Pires-daSilva, A. (2019). Liposome-based transfection enhances RNAi and CRISPR-mediated mutagenesis in non-model nematode systems. *Sci. Rep.* 9, 483. <https://doi.org/10.1038/s41598-018-37036-1>.
74. Moreno-Mateos, M.A., Vejnar, C.E., Beaudoin, J.D., Fernandez, J.P., Mis, E.K., Khokha, M.K., and Giraldez, A.J. (2015). CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* 12, 982–988. <https://doi.org/10.1038/nmeth.3543>.
75. Tuli, M.A., Daul, A., and Schedl, T. (2018). Caenorhabditis nomenclature. WormBook 2018, 1–14. <https://doi.org/10.1895/wormbook.1.183.1>.
76. Kramer, J.M., and Johnson, J.J. (1993). Analysis of mutations in the *sqt-1* and *rol-6* collagen genes of *Caenorhabditis elegans*. *Genetics* 135, 1035–1045. <https://doi.org/10.1093/genetics/135.4.1035>.
77. Schlager, B., Wang, X., Braach, G., and Sommer, R.J. (2009). Molecular cloning of a dominant roller mutant and establishment of DNA-mediated transformation in the nematode *Pristionchus pacificus*. *Genesis* 47, 300–304. <https://doi.org/10.1002/dvg.20499>.
78. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
79. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
80. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
81. Raney, B.J., Dreszer, T.R., Barber, G.P., Clawson, H., Fujita, P.A., Wang, T., Nguyen, N., Paten, B., Zweig, A.S., Karolchik, D., and Kent, W.J. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30, 1003–1005. <https://doi.org/10.1093/bioinformatics/btt637>.
82. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
83. Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295. <https://doi.org/10.1038/nbt.3122>.
84. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. <https://doi.org/10.1093/bioinformatics/btu638>.
85. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
86. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47. <https://doi.org/10.1093/nar/gkv007>.
87. Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Software* 25, 1–18. <https://doi.org/10.18637/jss.v025.i01>.
88. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
89. UniProt Consortium (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. <https://doi.org/10.1093/nar/gky1049>.
90. Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., Chen, W.J., Cho, J., Davis, P., Gao, S., Grove, C.A., Kishore, R., et al. (2020). WormBase: a modern Model Organism Information Resource. *Nucleic Acids Res.* 48, D762–D767. <https://doi.org/10.1093/nar/gkz920>.
91. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Palatin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
92. Bailey, T.L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
93. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. j.* 17, 10–12. <https://doi.org/10.14806/ej.17.1.200>.
94. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
95. Prijibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinformatics* 70, e102. <https://doi.org/10.1002/cbpi.102>.
96. Easton, A., Gao, S., Lawton, S.P., Bennuru, S., Khan, A., Dahlstrom, E., Oliveira, R.G., Kepha, S., Porcella, S.F., Webster, J., et al. (2020). Molecular evidence of hybridization between pig and human *Ascaris* indicates an interbred species complex infecting humans. *Elife* 9, e61562. <https://doi.org/10.7554/elife.61562>.
97. Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* 38, 3022–3027. <https://doi.org/10.1093/molbev/msab120>.
98. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. <https://doi.org/10.1101/gr.092759.109>.
99. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. <https://doi.org/10.1093/bioinformatics/btp033>.
100. Wagih, O. (2017). ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* 33, 3645–3647. <https://doi.org/10.1093/bioinformatics/btx469>.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
<i>E. coli</i> OP50	CGC	<a href="https://cgc.umn.edu/">https://cgc.umn.edu/</a>
<i>E. coli</i> NA22	CGC	<a href="https://cgc.umn.edu/">https://cgc.umn.edu/</a>
Chemicals, peptides, and recombinant proteins		
AsiSI	NEB	Cat# R0630
FseI	NEB	Cat# R0588
Exonuclease VII	NEB	Cat# R0630
Exonuclease T	NEB	Cat# M0625
Dynabeads™ MyOne™ Streptavidin C1	Invitrogen	Cat# 65001
USER Enzyme	NEB	Cat# M5505
Recombinant <i>S. pyogenes</i> Cas9 nuclease V3	IDT DNA	Cat# 1081058
Alt-R® CRISPR-Cas9 tracrRNA	IDT DNA	Cat# 1072533
TRIzol	Invitrogen	Cat# 15596026
Vectashield Plus antifade mounting media	Vector Laboratories	Cat# H-1900-10
Hoechst 33342	Invitrogen	Cat# H3570
Critical commercial assays		
ATAC-seq Kit	Active Motif	Cat# 53150
Ribocop rRNA Depletion Kit	LEXOGEN	Cat# 144
CORALL Total RNA-Seq Library Prep Kit	LEXOGEN	Cat# 146
Illumina DNA Preparation Kit	Illumina	Cat# 20018704
Deposited data		
Raw Genome Sequencing Reads	NCBI SRA	BioProject PRJNA882448; <a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA882448">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA882448</a>
Raw Data for RNA-seq, END-seq and ATAC-seq	NCBI GEO	GSE213886; <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213886">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE213886</a>
Experimental models: Organisms/strains		
<i>Oscheius tipulae</i> CEW1	CGC	<a href="https://cgc.umn.edu/">https://cgc.umn.edu/</a>
<i>Oscheius tipulae</i> PS2626 <i>him-1</i> mutant	CGC	<a href="https://cgc.umn.edu/">https://cgc.umn.edu/</a>
Wild isolates of <i>Oscheius tipulae</i>	[Baillé et al. <sup>52</sup> ]	<a href="https://www.justbio.com/tools/worms/search.php">https://www.justbio.com/tools/worms/search.php</a>
Oligonucleotides		
END-seq hairpin adaptor 1, 5'-Phos-GATCGGAAGAGCGTCGTGTAGGGAAAGA GTGTTU[Biotin-dT]U[Biotin-dT]UUACACTCTTCCGATC*T-3' [*phosphorothioate bond]	[Canela et al. <sup>48</sup> , Wong et al. <sup>49</sup> ]	N/A
END-seq hairpin adaptor 2, 5'-Phos-GATC GGAAGAGCACACGTCUUUUUUUUAGAC GTGTGCTCTCCGATC*T-3' [*phosphorothioate bond]	[Canela et al. <sup>48</sup> , Wong et al. <sup>49</sup> ]	N/A
DNA FISH probe: Telomere: 5'-Quasar 570-TTAGGCTTAGGCTTAGGCTTAGGC CRISPR gRNAs, primers, and END-seq Illumina TruSeq adaptors; see Table S1	Biosearch Technologies / [Wang et al. <sup>22</sup> ] This paper	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Software and algorithms</b>		
FIJI	[Schindelin et al. <sup>69</sup> ]	<a href="https://imagej.net/software/fiji/">https://imagej.net/software/fiji/</a>
CRISPRscan	[Moreno-Mateos et al. <sup>74</sup> ]	<a href="https://www.crisprscan.org">https://www.crisprscan.org</a>
Bowtie2 v2.4.2	[Langmead et al. <sup>78</sup> ]	<a href="https://github.com/BenLangmead/bowtie2">https://github.com/BenLangmead/bowtie2</a>
SAMtools v1.14	[Li et al. <sup>79</sup> ]	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
BEDTools v2.30.0	[Quinlan et al. <sup>80</sup> ]	<a href="https://github.com/arq5x/bedtools2">https://github.com/arq5x/bedtools2</a>
STAR	[Dobin et al. <sup>82</sup> ]	<a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>
StringTie	[Pertea et al. <sup>83</sup> ]	<a href="https://github.com/gperteau/stringtie">https://github.com/gperteau/stringtie</a>
HTSeq	[Anders et al. <sup>84</sup> ]	<a href="https://github.com/simon-anders/htseq">https://github.com/simon-anders/htseq</a>
DESeq2	[Love et al. <sup>85</sup> ]	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
BLAST	[Altschul et al. <sup>88</sup> ]	<a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
UniProt	[UniProt Consortium <sup>89</sup> ]	<a href="https://www.uniprot.org/">https://www.uniprot.org/</a>
WormBase	[Harris et al. <sup>90</sup> ]	<a href="https://wormbase.org/">https://wormbase.org/</a>
Pfam	[Mistry et al. <sup>91</sup> ]	<a href="https://www.ebi.ac.uk/interpro/entry/pfam/">https://www.ebi.ac.uk/interpro/entry/pfam/</a>
MEME	[Bailey et al. <sup>92</sup> ]	<a href="https://meme-suite.org/meme/">https://meme-suite.org/meme/</a>
Cutadapt v3.7	[Martin et al. <sup>93</sup> ]	<a href="https://github.com/marcelm/cutadapt">https://github.com/marcelm/cutadapt</a>
Trimmomatic	[Bolger et al. <sup>94</sup> ]	<a href="https://github.com/timflutre/trimmomatic">https://github.com/timflutre/trimmomatic</a>
SPAdes	[Prjibelski et al. <sup>95</sup> ]	<a href="https://github.com/ablab/spades">https://github.com/ablab/spades</a>
MEGA11	[Tamura et al. <sup>97</sup> ]	<a href="https://www.megasoftware.net/">https://www.megasoftware.net/</a>
Circos	[Krzywinski et al. <sup>98</sup> ]	<a href="http://circos.ca/">http://circos.ca/</a>
MetaMorph image acquisition software (version 7.7.10.0)	Molecular Devices, LLC	<a href="https://www.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metamorph-microscopy">https://www.moleculardevices.com/products/cellular-imaging-systems/acquisition-and-analysis-software/metamorph-microscopy</a>
Leica Application Suite X (LASX) (version 3.7.2.22383)	Leica	<a href="https://www.leica-microsystems.com/products/microscope-software/p/leica-las-x-ls/">https://www.leica-microsystems.com/products/microscope-software/p/leica-las-x-ls/</a>
<b>Other</b>		
M220 focused-ultrasonicator	Covaris	SKU: 500295
NovaSeq 6000	Illumina	N/A

**RESOURCE AVAILABILITY****Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Jianbin Wang ([jianbin.wang@utk.edu](mailto:jianbin.wang@utk.edu)).

**Materials availability**

Materials from this study are available on request.

**Data and code availability**

All sequencing data were deposited at the NCBI databases (SRA accession number: [PRJNA882448](https://www.ncbi.nlm.nih.gov/sra/PRJNA882448) and GEO accession number: GSE213886). The data for genome sequencing, gene models, RNA-seq, END-seq, and ATAC-seq are also available in a UCSC Genome Browser track data hubs that can be accessed with this link: <http://genome.ucsc.edu/s/jianbinwang/CEW1-genome-browser>. In addition, updated gene models, annotation, and RNA-seq datasets are available in <https://dnaelimination.utk.edu/protocols-data/>.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****Worm strains, culture and maintenance**

The CEW1 strain (originally isolated by Carlos E. Winter in Brazil) and the HIM-1 mutant (PS2626) were obtained from CGC (<https://cgc.umn.edu/>). The wild isolates of *O. tipulae* were from Dr. Marie-Anne Félix. The maintenance of *O. tipulae* is essentially the same as *C. elegans*. Briefly, *O. tipulae* strain CEW1, HIM-1 and other wild isolates were cultured using standard Nematode Growth Medium (NGM) plates seeded with *E. coli* OP50 bacteria.<sup>66</sup> To prepare a large number of synchronous worms, we used the *C. elegans* liquid culture protocol<sup>67</sup> with adaptations to promote dauer formation. Dauer worms were isolated using a 1% SDS solution followed by sucrose float. They were immediately plated on rich agar media containing an *E. coli* NA22 bacterial lawn. The dauer develop and mature into egg-laying worms at room temperature within ~36–40 hrs.

**METHOD DETAILS****Egg isolation, staged embryos and larvae, and other samples**

Egg-laying worms were washed off plates and passed through a 70-μm mesh. This step allows existing eggs to fall through the mesh, leaving the mature worms on the top. These worms were then bleached with sodium hypochlorite solution (Sigma Cat# 425044-250ML, 0.5 M NaOH, 3% bleach) to collect embryos (largely 2-4 cell stage, Figure S1). Early embryos were further developed in Virgin S Basal (VSB) medium (100 mM NaCl, 5.7 mM K<sub>2</sub>HPO<sub>4</sub>, 44.1 mM KH<sub>2</sub>PO<sub>4</sub>) at 25°C to reach to the desired embryonic stages (Figure S1). Briefly, embryos were staged by examining samples at half-hour intervals using a Leica DMIL-LED microscope with a 10X 0.25 NA lens and with a DMC6200 camera and LASX software. Images were analyzed by counting the number of cells visible in each embryo to determine the stage.

To prepare larval stages, the flow-through of the 70-μm mesh was further passed through a 25-μm mesh to separate mixed embryos from various stages of worms. These mixed embryos were developed in the VSB medium until all embryos reach to L1. The arrested L1s were plated on NGM with a robust OP50 lawn to develop at 20°C to synchronous stages of larvae, young adults, and mature adults. Alternatively, post-dauer larvae and worm samples were collected starting from the dauer harvested from the liquid culture. For adult males, we handpicked 200 males from the HIM-1 mutant for RNA-seq library replicate.

**Hoechst staining**

The staining procedure was adapted from a previous study on *Ascaris* embryos.<sup>68</sup> Briefly, embryos obtained from bleached adults were spun down at 1,000 x g for 2 minutes in a 1.5 mL Eppendorf tube and resuspended in 500 μL of fixative solution containing 50% methanol (Fisher Chemical, A412SK-4) and 2% para-formaldehyde (Ted Pella, Inc., Cat # 18505) in 1X PBS buffer. The embryos were freeze-cracked by submerging the tube in liquid nitrogen until the solution was completely frozen, followed by thawing in a beaker of lukewarm water. This process was repeated at least five times to ensure thorough eggshell cracking. The tube was opened between each round to release trapped nitrogen gas. After spinning and removing the fixative, the embryos were rehydrated by incubation in 25% methanol in 1X PBS for one minute, followed by incubation in 1X PBS for one minute. Embryos were then incubated with 500 μL of Hoechst 33342 (1 μg/mL) (Invitrogen, Fisher Cat# H3570) dye at room temperature for ten minutes in the dark with rotation. Embryos were washed in blocking solution and mixed with Vectashield plus antifade mounting media (Vector Laboratories, H-1900-10) and pipetted onto a glass slide. For each slide, a coverslip was placed on the sample and the slide was wrapped in a paper towel, turned upside-down, and gently flattened with a large book for 3 minutes. Coverslips were sealed using clear nail polish (Ted Pella, Inc., Fisher Cat# NC1849418). Imaging was performed on a Nikon Eclipse Ti inverted spinning disk confocal microscope using a 100X 1.49 NA oil immersion objective lens. Images were collected on a Photometrics Evolve 512 EMCCD camera and acquired using MetaMorph software (version 7.7.10.0, Molecular Devices, LLC). Images were acquired as Z-stacks using a step size of 0.3 μm. Image processing was performed using FIJI.<sup>69</sup> All contrast adjustments in figures are linear.

**DNA FISH**

Staged embryos were fixed and subjected to freeze-cracking as described above. Embryos were treated with 100 μg/mL of RNase A (Invitrogen, AM2271) at 37°C for 30 minutes with rotation. Embryos were washed in 1X PBS, then incubated in FX-signal enhancer (Invitrogen, I36933) at room temperature for 30 minutes with rotation. Embryos were resuspended and incubated in blocking solution (1% BSA in 1X PBS) at room temperature for one hour with rotation. Samples were then incubated in pre-hybridization buffer (10% formamide [Fisher Scientific, Cat# 4650] in 2X SSC buffer [Fisher Scientific, Cat# AM9770]) at 37°C for five minutes with rotation. Embryos were next incubated in pre-heated denaturation buffer (50% formamide in 2X SSC buffer) at 75°C for five minutes. The telomere-Quasar570 probe (TTAGGCTTAGGCTTAGGCTAGGC, from Biosearch Technologies) was added to hybridization buffer and preheated at 75°C. Embryos were incubated in the hybridization buffer overnight at 37°C in the dark with rotation. An equal volume of 2X SSC buffer was then added to the samples, which were next spun down and washed with 0.4 X SSC buffer at 45°C for 15 minutes followed by a wash in 2X SSC buffer, 0.1% NP-40 alternative (EMD Millipore, Cat# 492016) at room temperature for 15 minutes. Samples were stained in Hoechst-33342, mounted to slides, and imaged as described above.

### RNA isolation and sequencing

Total RNA was prepared using TRIzol (Invitrogen Cat# 15596026) protocol. The quality of the RNA was evaluated with the TapeStation 4200 (Agilent) and quantified with the Qubit™ 4 Fluorometer (Thermo Fisher). The ribosomal RNAs were removed using the RiboCup rRNA Depletion Kit (LEXOGEN Cat# No 144). The RNA-seq libraries were constructed using the CORALL Total RNA-Seq Library Prep Kit (LEXOGEN Cat# No 146) and sequenced using an Illumina NovaSeq 6000 at the University of Colorado Anschutz Medical Campus Genomics Core.

### Genomic DNA extraction and sequencing

Cultures of *O. tipulae* were grown on 150 mm plates of rich agar seeded with *E. coli* NA22. When the bacterial food source was near exhaustion, worms were washed off plates with M9 buffer and centrifuged at 170 x g for 1 min. The pelleted worms were then purified from microbial contaminants using a 35% sucrose floatation.<sup>70</sup> Genomic DNA was extracted from harvested worms using Genomic-tips (Qiagen Cat# 10223). The resulting high molecular weight genomic DNA was used to prepare sequencing libraries using the Illumina DNA Prep kit (Cat# 20018704) following the manufacturer's instructions. The libraries were sequenced using an Illumina NovaSeq 6000. The 23 wild isolates were sequenced using the same procedure.

### CRISPR-Cas9 modification of the *O. tipulae* genome

We adapted CRISPR-Cas9 procedures from *C. elegans*<sup>71,72</sup> and other nematodes<sup>73</sup> and developed a protocol for *O. tipulae*. All CRISPR-Cas9 reagents (crRNAs, tracrRNAs, Cas9, Cat# 1081058) were obtained from IDT DNA (Coralville, IA). Locus-specific crRNAs were selected using a combination of rating algorithms from IDT DNA and CRISPRscan.<sup>74</sup> Candidate crRNAs that scored at least 40 (range 0-100) in both algorithms were considered for further experimentation. A list of all crRNAs and repair templates used is provided in Table S1. The injection mixes were synthesized as follows: A CRISPR-Cas9 RNP to induce a dominant roller mutation in *O. tipulae* *rol-6* (see below) for co-CRISPR selection was made by mixing 0.5 µL 10 mg/mL Cas9 protein into 4.5 µL IDT duplex buffer. 2.5 µL of 100 µM tracrRNA was added, followed by 2.5 µL of 100 µM *rol-6* crRNA. A locus-specific CRISPR-Cas9 RNP was produced by mixing 0.5 µL 10 mg/mL Cas9 protein into 3.5 µL IDT duplex buffer. We then added 4.0 µL of 100 µM tracrRNA, followed by 1.0 µL of 200 µM of each crRNA that flanked the elimination motif. Both CRISPR-Cas9 RNP complexes were incubated at 37°C for 15 minutes, then 1.0 µL of 100 µM repair template was added to each RNP mix to mediate homology-directed repair of a roller phenotype or mutation of an SFE motif. The two RNP complexes were mixed and centrifuged at 10,000 x g for two minutes. Lipofectamine (Invitrogen Cat# 13778030) was added to 3% v/v to the RNP-containing supernatant, and the mix was incubated at room temperature for 20 minutes prior to loading needles for injection.

Injections were done using a Nikon Eclipse Ti-U inverted scope with DIC optics and a Narishige IM 300 injector with an input pressure of 70 psi and an injection pressure of 20 psi. Injection needles were from World Precision Instruments Inc. (cat. # 1B100F-4) and were pulled on a Sutter P-87 puller using a two-step cycle of heat 330, pull 0, velocity 20, time 200, followed by heat 330, pull 20, velocity 65, time 150. The pressure was set to 400.

Well-fed young adult (4 days old) *O. tipulae* were picked to pads of dried 2% agarose and covered with 700 weight halocarbon oil (Sigma Cat# 9002839) for injection. Injected worms were released from agarose pads by gentle resuspension in M9 buffer, picked to OP50-seeded NGM plates, and incubated at 25°C. Plates were screened for F1 worms with a roller phenotype at 3-5 days post-injection. In our hands, as many as 15% of injected P0 worms produced one or more F1 progeny with a roller phenotype. F1 rollers were allowed to produce F2 progeny, and either the parental F1 roller or a pool of 10 F2 progeny were screened by PCR<sup>71</sup> to detect the CRISPR-Cas9-induced edit of interest. Potential mutants were plated as single F2s, and their progeny were screened for homozygosity for the mutation of interest via PCR.

### Mutant strains nomenclature

We named the *O. tipulae* mutants using the *C. elegans* nomenclature.<sup>75</sup> Briefly, our laboratory allele designation "pde" was followed by Df (deficiency) or Dp (Duplication or insertion) and the site of the mutation. For example, the deletion of the 61 bp sequences in the SFE site from the left end of chromosome II is named as *pdeDf61[sfe-2L]*. Similarly, the insertion of the 22 bp sequences in the SFE site from the right end of chromosome X is named as *pdeDp22[sfe-XR]*.

### Identification of a *rol-6* orthologue in *O. tipulae*

Certain alleles of the *C. elegans* *rol-6* locus, which encodes a cuticular collagen, give rise to a dominant "roller" phenotype<sup>76</sup> that is readily scored and can be used for co-CRISPR selection. Orthologous *rol-6* genes have also been identified in *Pristionchus pacificus* and *Auanema* species.<sup>73,77</sup> Using the *Auanema rhodensis* ROL-6 protein sequence, we scanned the *O. tipulae* genome and identified a candidate *O. tipulae* *rol-6* gene (Figure S7), encoding a protein that has 82% identity through an 88-amino acid segment of *A. rhodensis* ROL-6 that encompasses the critical R→C substitution associated with a roller phenotype (Figure S7). The introduction of the R→C substitution via CRISPR-Cas9 and a repair template harboring the desired base substitutions (Figure S7) resulted in *O. tipulae* worms with a roller phenotype essentially indistinguishable from that of *C. elegans* with a *su1006* allele of the *rol-6* gene. This mutant *rol-6* repair template was used in all subsequent experiments as a co-CRISPR selection marker.

### END-seq library preparation

Staged embryos were used to create END-seq libraries using an adapted protocol.<sup>48,49</sup> Briefly, embryos were embedded in agarose plugs to protect the DNA from exogenous breaks. Some plugs were digested with the restriction enzyme AsISI (NEB, catalog # R0630) or FseI (NEB, catalog # R0588) to generate DSBs as an internal control. DSBs were blunted with exonuclease VII (NEB, catalog # R0630) and exonuclease T (NEB, catalog # M0625). Blunt ends were A-tailed and capped with END-seq adaptor 1, a biotinylated hairpin adaptor. DNA was liberated from the plugs and sheared to ~200–300 bp with a Covaris M220 focused-ultrasonicator (130 µL tube, peak power 50, duty 16, cycles/burst 200 for 420 seconds). DNA fragments containing END-seq adaptor 1 were isolated with Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen, catalog # 65001). The other ends broken by sonication were repaired and A-tailed with END-seq adaptor 2. Hairpins were digested with USER (NEB, catalog # M5505), and the DNA was amplified with Illumina TruSeq primers and barcodes. The libraries were sequenced with NovaSeq 6000.

### ATAC-seq library preparation

Staged embryos were used to carry out ATAC-seq. Embryos were resuspended in 5 mL of cold ATAC-seq lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 0.1% Triton-X 100, 0.1% NP40 alternative) and transferred to a dounce homogenizer kept cold on ice. Embryos were dounced ten times before being transferred to a 15 mL Eppendorf tube and spun down at 500 × g for 5 minutes at 4°C. Samples were fragmented and subsequently purified using an ATAC-seq kit (Active Motif, Cat# 53150) according to the cell sample protocol. Libraries were amplified through ten rounds of PCR and sequenced using NovaSeq 6000.

### Bioinformatic data analysis

#### Genome mapping

Genomic reads from staged embryos, SFE mutants, and wild isolates were mapped to the *O. tipulae* telomere-to-telomere genome<sup>34</sup> using bowtie2<sup>78</sup> and SAMtools<sup>79</sup> to generate BAM files. The read coverage was obtained using BEDTools<sup>80</sup> to produce a BEDGRAPH file that was loaded into UCSC genome browser track data hubs.<sup>81</sup>

#### RNA-seq data analysis

Ribosomal RNA reads were filtered out using Bowtie2.<sup>78</sup> The non-rRNA reads from each sample were mapped to the *O. tipulae* genome with STAR.<sup>82</sup> The transcripts for each sample were assembled separately using the StringTie<sup>83</sup> and merged into a non-redundant set of transcripts. The predicted coding regions (<https://github.com/TransDecoder>) from the transcripts were used to further evaluate and to remove artifacts, cryptic transcripts, and transcriptional noises based on their coding capacity (< 50%), the lack of splicing, and/or low expression. For expression analysis, we used HTSeq<sup>84</sup> to get the read count for all transcripts. The DESeq2<sup>85</sup> package was used to determine differentially expressed genes between the developmental stages with a fold-change cutoff of 2 and an adjusted p-value ≤ 0.05. Several R packages were also used to generate the heatmap,<sup>86</sup> the volcano plot<sup>87</sup> and the PCA plot (<https://rdrri.io/cran/factoextra/>).

#### Gene annotation and functional analysis

The transcripts were used to BLAST<sup>88</sup> against protein databases (NCBI nr, UniProt, <sup>89</sup> Swiss-Prot, WormBase<sup>90</sup> *C. elegans* and *Ascaris*) to assign annotation. The translated protein sequences from TransDecoder were used to search against the Pfam database<sup>91</sup> (hmmscan E-value cutoff <=10<sup>-10</sup>). Genes that matched to *C. elegans* (WBgene ID) were used for Gene Ontology and tissue enrichment analysis.<sup>45,46</sup>

#### Motif analysis

*De novo* motif identification was initially done by analyzing 1 kb regions from the 12 chromosomal breakage sites in CEW1. These sequences were analyzed with MEME<sup>92</sup> a tool to discover ungapped motifs. *De novo* identification of the motif sequence was reinforced by including 23 wild stains in the analysis. The MEME motif matrix was used as input for FIMO<sup>53</sup> to search for alternative sites within the CEW1 strain and among 23 wild isolates using default parameters (p-value < 1.0E-4). Telomeric sequences were filtered from the FIMO-generated motifs list, and a cut-off score of 50 was initially used to call the alternative motif sites.

#### END-seq data analysis

Reads (R1 file only) were mapped to the *O. tipulae* reference genome<sup>34</sup> with Bowtie2<sup>78</sup> and processed with SAMtools.<sup>79</sup> The 5' position of each read was mapped, and separated by strand using BEDTools<sup>80</sup> (genomecov). For comparison among libraries, the reads were normalized to one million genome-mapped reads. Telomere-unique reads were (R1 and R2 files) defined as those containing two or more sequential telomeric repeats (TTAGGCTTAGGC). Germline genomic regions that contain telomere repeats were filtered out. All telomere-unique reads were converted to the G-rich strand (TTAGGC). Cutadapt<sup>93</sup> (-j 0 -m 20 -b TTAGGCTTAGGC) was then used to trim the telomeric sequences. These trimmed reads were mapped to the genome, and the telomere-unique junctions were used to identify the telomere addition sites.

#### ATAC-seq data analysis

ATAC-seq reads were mapped to the *Oscheius tipulae* genome using bowtie2<sup>78</sup> and processed using SAMtools<sup>79</sup> and BEDTools.<sup>80</sup> The reads were normalized based on the genome-mapped reads. ATAC-seq data was uploaded to the UCSC genome browser track data hubs.<sup>81</sup>

**Wild isolates sequence analysis**

Raw reads were preprocessed with Trimmomatic<sup>94</sup> to remove potential adapters and trim low-quality nucleotides. *De novo* genome assemblies of 23 wild strains were done using SPAdes<sup>95</sup> with “–careful” option recommended for small genomes that reduces the number of mismatches and short indels. The mitochondrial genomes were assembled using a reference-assisted *de novo* approach as described.<sup>96</sup> The phylogenetic tree for the mitochondrial genomes was constructed using MEGA11.<sup>97</sup> The genomic reads were mapped to the CEW1 genome as described in the genome mapping section. Telo-mere-unique reads, described in the END-seq data analysis, were mapped to the CEW1 genome to identify potential sites of alternative SFEs.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Statistical parameters, including sample numbers, mean and standard deviation or error, were included in Figures, Supplemental Figures and the Figure legends. For the statistical analyses of the data shown in [Figure S4](#) and [Tables S4](#), p values were computed using the Wald test; to control for false positives, p values were adjusted using DESeq2’s built-in Benjamini-Hochberg correction for multiple hypothesis testing. An adjusted p value < 0.05 is considered as statistically significant.