Degradation of the Repetitive Genomic Landscape in a Close Relative of *Caenorhabditis elegans*

Gavin C. Woodruff *, [1] and Anastasia A. Teterina [1,2]

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR

[2]Center of Parasitology, Severtsov Institute of Ecology and Evolution RAS, Moscow, Russia

***Corresponding author**: E-mail: gavincw@uoregon.edu.

Associate editor: Jian Lu

Genome sequences and annotations were retrieved from WormBase ParaSite (parasite.wormbase.org) and the *Caenorhabditis* genomes project (caenorhabditis.org).

Abstract

The abundance, diversity, and genomic distribution of repetitive elements is highly variable among species. These patterns are thought to be driven in part by reproductive mode and the interaction of selection and recombination, and recombination rates typically vary by chromosomal position. In the nematode *Caenorhabditis elegans*, repetitive elements are enriched at chromosome arms and depleted on centers, and this mirrors the chromosomal distributions of other genomic features such as recombination rate. How conserved is this genomic landscape of repeats, and what evolutionary forces maintain it? To address this, we compared the genomic organization of repetitive elements across five *Caenorhabditis* species with chromosome-level assemblies. As previously reported, repeat content is enriched on chromosome arms in most *Caenorhabditis* species, and no obvious patterns of repeat content associated with reproductive mode were observed. However, the fig-associated *C. inopinata* has experienced repetitive element expansion and reveals no association of global repeat density with chromosome position. Patterns of repeat superfamily specific distributions reveal this global pattern is driven largely by a few repeat superfamilies that in *C. inopinata* have expanded in number and have weak associations with chromosome position. Additionally, 15% of predicted protein-coding genes in *C. inopinata* align to transposon-related proteins. When these are excluded, *C. inopinata* has no enrichment of genes in chromosome centers, in contrast to its close relatives who all have such clusters. Forward evolutionary simulations reveal that chromosomal heterogeneity in recombination rate alone can generate structured repetitive genomic landscapes when insertions are weakly deleterious, whereas chromosomal heterogeneity in the fitness effects of transposon insertion can promote such landscapes across a variety of evolutionary scenarios. Thus, patterns of gene density along chromosomes likely contribute to global repetitive landscapes in this group, although other historical or genomic factors are needed to explain the idiosyncrasy of genomic organization of various transposable element taxa within *C. inopinata*. Taken together, these results highlight the power of comparative genomics and evolutionary simulations in testing hypotheses regarding the causes of genome organization.

Key words: **transposable elements, genome organization, recombination, *Caenorhabditis*.**

Introduction

Repetitive elements are a conspicuous feature of eukaryotic genomes. Over half of the human genome comprised such elements (de Koning et al. 2011), and the maize genome has a repeat content of over 80% (Baucom et al. 2009; Schnable et al. 2009). But at the same time, the range in repeat content among eukaryotic genomes is great, with some species having a scant number of repetitive elements (0.8% in one species of bdelloid rotifer; Nowell et al. 2018). And not only does the global repeat content among genomes vary—heterogeneity in repeat content both within and between chromosomes occurs (Rizzon et al. 2002; Stitzer et al. 2019). Furthermore, repeat density in genomes has been observed to covary with patterns of genomic diversity (Clark et al. 2007), recombination rate (Rizzon et al. 2002), gene density (Medstrand et al. 2002), chromatin state (Peacock et al. 1978; Verma 1988; Peng and Karpen 2008), centromeric regions (Plohl et al. 2014), and physical spatial position (Guelen et al. 2008). Repetitive elements are then a major feature of genomic organization, and the origin and maintenance of their genomic landscape demands explanation.

Transposable elements are generally considered deleterious by replicating at the expense of its host and abrogating functional sites through insertion. This is largely consistent with experiments revealing that fitness declines with increased transposable element activity (Pasyukova et al. 2004; Bégin and Schoen 2006). Thus, it has been proposed that variation in repeat content among species is driven by variation in population size; weaker selection in species with smaller population sizes should lead to increased repeat content (Lynch 2007). This is also thought to explain within-genome heterogeneity in repeat content—low-recombining regions have higher repeat content than regions with high recombination rates in multiple systems (including yeast, Pan

Article

et al. 2011; mice, Jensen-Seaman et al. 2004; humans, Jensen-Seaman et al. 2004; *Arabidopsis thaliana*, Wright et al. 2003; Lockton et al. 2008; Lockton and Gaut 2010; Dubin et al. 2015; Stuart et al. 2016; Kent et al. 2017; and maize, Liu et al. 2009, among others). The dynamics of transposable element evolution and their interaction with recombination have long been studied in *Drosophila*; in this system, transposable elements are generally enriched in regions of low recombination (Charlesworth and Langley 1989; Bartolomé et al. 2002; Blumenstiel et al. 2002; Rizzon et al. 2002; Dolgin and Charlesworth 2008; Lee and Langley 2010; Petrov et al. 2011; Comeron et al. 2012; Cridland et al. 2013; Barrón et al. 2014; Kofler et al. 2015; Stuart et al. 2016; Adrion et al. 2017; Kent et al. 2017). This is presumably because of weakened selection in low-recombining regions (Hill and Robertson 1966; Charlesworth and Langley 1989; Lee and Langley 2010; Barrón et al. 2014). In mice and humans, the association of transposable element content with recombination rate varies depending on repeat type (Myers et al. 2005; Shifman et al. 2006). However, transposable element activity was not correlated with recombination rate after accounting for chromatin states in *Drosophila melanogaster* (Adrion et al. 2017). Others have noted that transposable element abundance should either increase or decrease depending on the model of natural selection and the mode of reproduction (Wright and Schoen 1999; Bestor 2000; Morgan 2001; Glémin et al. 2019). At the same time, it has also been proposed that repetitive elements themselves can be adaptive (Shapiro and Von Sternberg 2005), and individual cases of adaptation via transposable element insertion abound (Oliver and Greene 2009; Casacuberta and González 2013). And in nematodes, repeat content is atypically positively correlated with recombination (Duret et al. 2000), and little difference is observed between selfing and outcrossing close relatives (Fierst et al. 2015). Thus, it remains unclear to what degree selection, drift, recombination, and reproductive mode contribute to the evolution of repetitive genomic organization.

One approach toward understanding this problem is investigating a group of recently diverged species, where evolutionary signals are still detectable (Jenner and Wills 2007; Raff 2012). The nematode *Caenorhabditis elegans* was the first metazoan to have its genome sequenced, and consequently, it is among the more thoroughly annotated and understood sequences available (Gerstein et al. 2010). About 12–16% of the *C. elegans* genome consists of transposable elements (*C. elegans* Sequencing Consortium 1998; Sijen and Plasterk 2003; Laricchia et al. 2017), and active elements are dominated by the Tc1/mariner superfamily of DNA transposons (Eide and Anderson 1985; Bessereau 2006). Intrachromosomal heterogeneity of numerous genomic features has long been known in this species, which has led to the definition of high-recombining, repeat-rich chromosome "arms" and low-recombining, gene-rich chromosome "centers" or "clusters" (Barnes et al. 1995; Cutter et al. 2009). These patterns appear to not be influenced by discrete centromeres as *C. elegans* has holocentric chromosomes, where spindle attachment sites span the entire length of chromosomes (Albertson and Thomson 1982; Howe et al. 2001). Previous studies have shown similar clusters of protein-coding genes in other members of the genus (Stein et al. 2003; Kanzaki et al. 2018; Yin et al. 2018; Teterina et al. 2020). Furthermore, recombination maps of *C. briggsae* reveal that its intrachromosomal variation in recombination rate is conserved between this species and *C. elegans* (Rockman and Kruglyak 2009; Ross et al. 2011). These also reveal that recombination domain organization of the X chromosome is comparable with that of autosomes, although recombination rates vary both within and between chromosomes and their domains (Rockman and Kruglyak 2009; Ross et al. 2011). Regardless, many genomes of close relatives of *C. elegans* have recently been sequenced, some of which have been assembled to chromosome-level contiguity (Fierst et al. 2015; Kanzaki et al. 2018; Ren et al. 2018; Yin et al. 2018; Stevens et al. 2019; Teterina et al. 2020). Here, we harness these new resources to interrogate the evolution of repetitive genomic landscapes among five *Caenorhabditis* species. We find a conserved chromosomal distribution of repetitive elements among four species, whereas the ecologically and morphologically divergent *C. inopinata* (Kanzaki et al. 2018; Woodruff and Phillips 2018; Woodruff et al. 2018) harbors an atypically uniform repetitive landscape driven by a handful of transposable element superfamilies.

Results

Repeat Density Covaries with Chromosomal Position in All Species but *C. inopinata*

The genomic landscape of repetitive elements was inferred in five *Caenorhabditis* assemblies (fig. 1) through a combination of de novo and element class-specific methods (see Materials and Methods; fig. 2). As previously described in nematodes (Duret et al. 2000; Rizzon et al. 2003; Stein et al. 2003; Cutter et al. 2009; Yin et al. 2018), four of the assemblies reveal an enrichment of repetitive elements on chromosome arms relative to chromosome centers (fig. 2; supplementary figs. 3 and 4, Supplementary Material online). Although there is a range in the repeat density difference between chromosome arms and centers among species (1–18% repeat content/10-kb window), *C. remanei* reveals the greatest difference in mean repeat content (fig. 2; supplementary figs. 3 and 4, Supplementary Material online). Surprisingly, the closest relative of *C. elegans* used in this study, *C. inopinata*, revealed far less enrichment in chromosome arms relative to chromosome centers (1% arms/centers difference; fig. 2; supplementary figs. 3 and 4, Supplementary Material online). After normalizing all genomic positions relative to chromosome centers, *C. inopinata* is the only species that did not reveal a significant relationship between chromosome position and repeat density in a linear model (supplementary figs. 4 and 5, Supplementary Material online; $r^2 = 1.8 \times 10^{-5}$; $F = 1.2$; $\beta_1 = -1.5$; $P = 0.27$). Conversely, all other species had a positive relationship between repeat density and distance from chromosome center (supplementary figs. 4 and 5, Supplementary Material online; $r^2 = 0.078$–0.24; $F = 999$–$3{,}165$; $\beta_1 = 32$–60; $P < 0.0001$ for all). There is then a largely

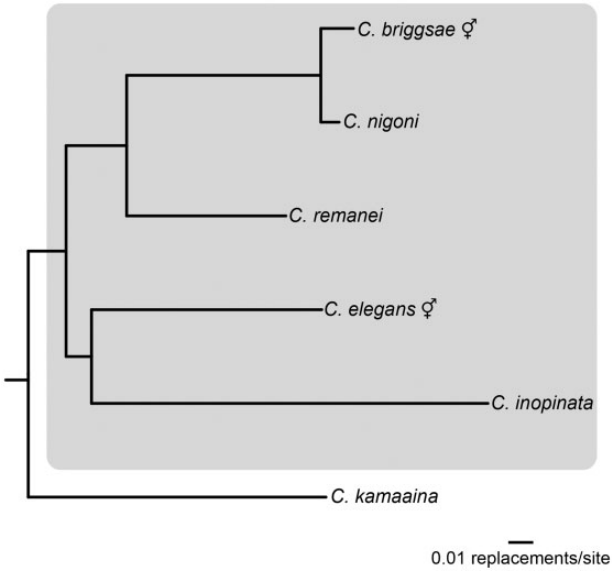

Fig. 1. *Caenorhabditis* phylogeny. Species addressed in this study are included in the gray box. *Caenorhabditis elegans* and *C. briggsae* are hermaphroditic species (indicated by the ⚥ symbol), whereas *C. nigoni*, *C. remanei*, and *C. inopinata* are male/female. The topology is derived from the Bayesian phylogeny inferred from protein sequences in (Stevens et al. 2019).

conserved pattern of enrichment of repetitive elements in chromosome arms although it varies in its extent across species; *C. inopinata* is an exception in that it has almost no such detectable global repeat chromosomal organization.

Divergent Genomic Repetitive Landscapes Are Driven by Diversity in Repeat Taxon Abundance and Chromosomal Distributions

Transposable elements harbor abundant structural and replicative diversity. To understand the impact of such diversity on the genomic repetitive landscape, transposable elements were systematized into a taxonomy (supplementary_tables.xls, sheet 23, Supplementary Material online) informed by (Wicker et al. 2007). The genomic landscapes of repeat taxa vary widely (fig. 3; see supplementary_figures_index.docx, Supplementary Material online, for links to the genomic distributions of all repeat taxa in all species). Repeat superfamilies (of which 26 were found among the repetitive elements in *Caenorhabditis*) can reveal conserved genomic landscapes with enrichment on chromosome arms (hAT superfamily, fig. 3) or with little apparent chromosomal distribution (Mutator superfamily, fig. 3). Conversely, repeat superfamilies can vary widely in their genomic landscapes among species (such as PiggyBac, Bel-Pao, Tc1-Mariner, and RetroTransposon-like Element (RTE) superfamilies; fig. 3). Genomic repetitive landscapes are thus composed of dozens of repeat taxa that can each harbor idiosyncratic abundances and chromosomal landscapes among species and within genomes.

To further explore genomic landscapes of repeat taxa and their contributions to global repetitive landscapes, repeat taxon densities among chromosome arms and centers were compared (fig. 4*a* and *b*; supplementary figs. 6–12, Supplementary Material online). Additionally, the total genomic abundances of various repeat taxa were also compared (fig. 4*c* and *d*; supplementary figs. 6–12, Supplementary Material online). Among repeat classes (here including unclassified repeats, satellite DNA, and low-complexity repetitive elements), transposable elements were most abundant (supplementary fig. 4, Supplementary Material online), with class II DNA transposons generally being the dominant repeat class (supplementary figs. 6 and 7, Supplementary Material online). Consistent with global repeat distributions, transposable elements tend to be enriched on chromosome arms (supplementary fig. 6, Supplementary Material online). However, in *C. inopinata*, class II DNA transposons are uniformly distributed along the chromosomes (arm-center Cohen's *d* effect size = 0.044, linear model $P = 0.69$; supplementary figs. 6 and 7, Supplementary Material online), and class I DNA retrotransposons are enriched on chromosome centers, in exception to all other species (arm-center Cohen's *d* effect size = −0.18, linear model $P < 0.0001$; supplementary figs. 6 and 7, Supplementary Material online). Thus, variation in the chromosomal distributions and abundances of repeat classes underlies diversity in global repeat content.

Repeat taxa of lower ranks were interrogated further to understand the exceptional repetitive landscape of *C. inopinata*. All species but *C. inopinata* reveal a significant correlation between repeat superfamily abundance and arm-center effect size (fig. 4*c*; supplementary fig. 13, Supplementary Material online; *C. inopinata*: $r^2 = 0.061$, $F = 1.2$, $\beta_1 = -1.6$, $P = 0.29$; all other species: $r^2 = 0.21$–0.64, $F = 5.0$–38, $\beta_1 = 2.1$–4.5, $P = 3.4 \times 10^{-6}$–0.038). Some superfamilies that reveal atypical chromosomal distributions are highly abundant in the *C. inopinata* genome, particularly Tc1-Mariner (class II transposon; 10.2% of the genome), RTE (class I LINE retrotransposon; 5.7%), Bel-Pao (class I long terminal repeat [LTR] retrotransposon; 2.2%), and Gypsy (class I LTR retrotransposon; 2.0%; fig. 4*c* and *d*), consistent with previous descriptions of the *C. inopinata* genome (Kanzaki et al. 2018). When these four superfamilies are excluded, the global repetitive landscape retains a chromosomal distribution that resembles those of its close relatives ($r^2 = 0.048$, $F = 621$, $\beta_1 = 17.7$, $P < 0.0001$; fig. 5). This is also consistent with global genomic distributions of discrete repeat counts and repeat lengths (see supplementary_figures_index.docx, Supplementary Material online, for links to these genomic landscapes), which reveal that repetitive regions are longer in chromosome centers in *C. inopinata*. Thus, divergence in the repetitive genomic landscape can be driven in large part by the activity of a small number of repeat superfamilies that nonetheless differ greatly in chromosomal distribution and mode of replication.

Gene Density Is Negatively Correlated with Repeat Content in All Species but *C. inopinata*

Previous reports have noted that the genomic landscape of gene density mirrors that of repeat content in nematodes, and genes are enriched in chromosome centers relative to arms (Cutter et al. 2009; Fierst et al. 2015; Yin et al. 2018).

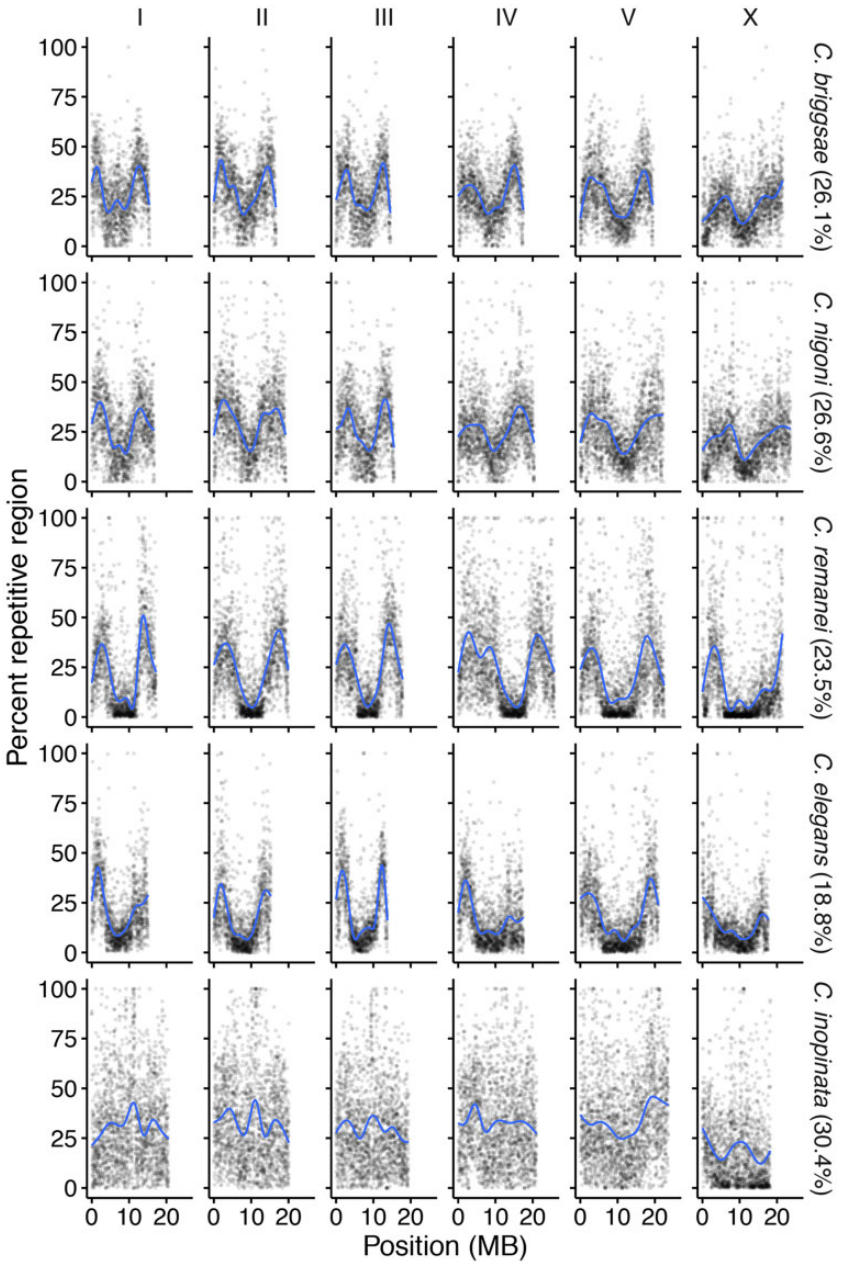

Fig. 2. The genomic landscape of repetitive elements in five *Caenorhabditis* assemblies. Columns represent the six chromosomes; rows are the species ordered phylogenetically as in figure 1. Plotted are the percentages of 10-kb windows that contain repetitive regions by genomic position. Percentage of the whole genome that is repetitive is noted in parentheses. Blue lines were fit with generalized additive models.

Nonrandom distributions of genes and repetitive elements also overlap with domains of high and low recombination. To explore if gene density is associated with repeat content in the assemblies addressed here, gene densities were estimated with publicly available genome annotations. When paired with estimates of repeat content, we found that, as expected, genes are moderately enriched in chromosome centers relative to arms in all five species (supplementary figs. 14 and 15, Supplementary Material online; $r^2 = 0.025–0.12$, $F = 33–180$, $\beta_1 = -6.2$ to -0.20, $P < 0.0001$ for all) and that gene density is negatively correlated with repeat content in four assemblies (fig. 6; $r^2 = 0.059–0.16$, $F = 66–235$, $\beta_1 = -0.73$ to -0.35, $P < 0.0001$ for all but *C. inopinata*). This pattern is largely driven by differences between arm-center chromosome domains, although the relationship persists within chromosome arms and centers (except in *C. elegans* chromosome centers; supplementary figs. 16 and 17, Supplementary Material online). However, *C. inopinata* revealed a positive correlation among repeat content and gene density (fig. 6; $r^2 = 0.050$, $F = 65$, $\beta_1 = 0.54$, $P < 0.0001$). As the previous description of the *C. inopinata* genome mentioned a transposable element insertion within the highly conserved sex determination protein-coding gene *her-1* (Kanzaki et al. 2018), we reasoned this positive relationship may be due to the presence of repetitive content in predicted protein-coding genes. To test this, we aligned all predicted protein-

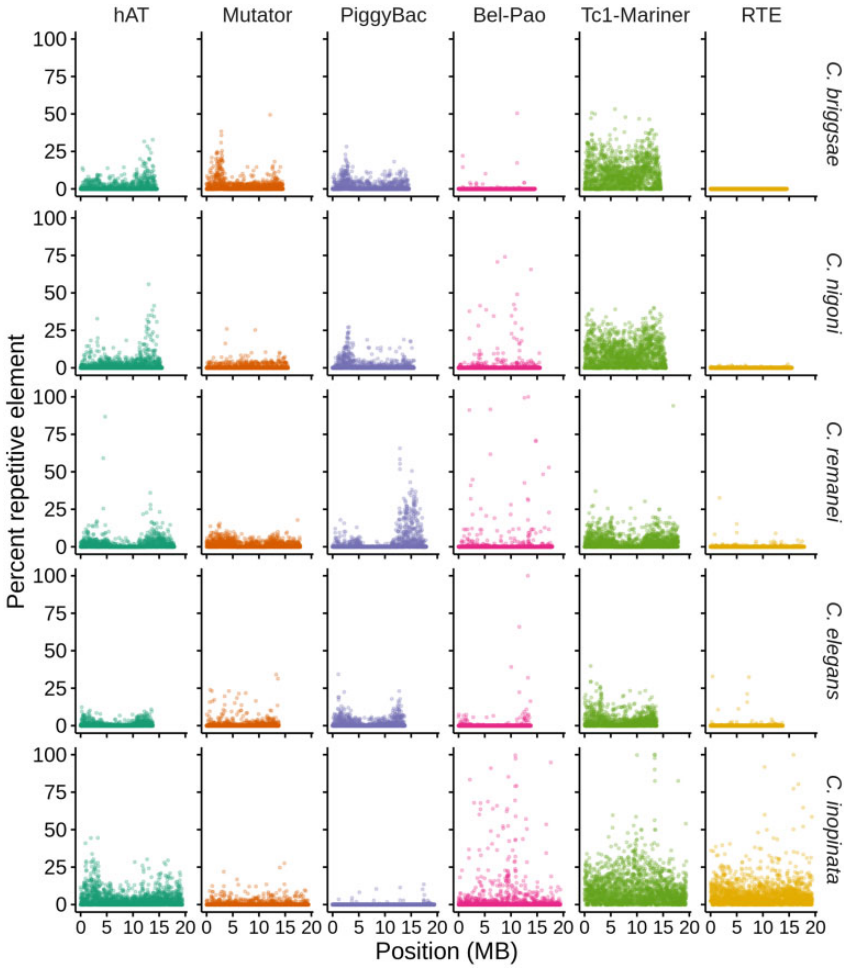

Fig. 3. Repeat superfamilies vary in their chromosomal distributions. Plotted are the percentages of 10-kb windows that contain a given repeat superfamily (columns) along chromosome III in five species (rows). Six superfamilies among 26 detected were chosen to illustrate the diversity in repetitive landscapes (see supplemental_figures_index.docx for links to all distributions of all repeat taxa on all chromosomes, Supplementary Materia online).

coding genes from 28 *Caenorahbditis* genomes and one outgroup (*Diploscapter coronatus*; Hiraki et al. 2017) to the TransposonPSI transposon protein database (transposonpsi.sourceforge.net, last accessed May 7, 2020.). In most *Caenorhabditis* genomes, only a small fraction of predicted protein-coding genes aligns to transposon-related proteins (median = 1.08%; interquartile range = 0.47%). However, in *C. inopinata* and *C. japonica*, a substantial fraction of their protein sets aligns to such proteins (supplementary fig. 18, Supplementary Material online; *C. inopinata*, 15%; *C. japonica*, 12%). Within *C. inopinata*, of the 3,349 proteins that aligned to the transposon database, 860 also aligned to *C. elegans* nonrepetitive proteins. Indeed, after *C. inopinata* proteins aligning exclusively to transposable elements are removed, there is no significant relationship between repeat content and gene density (fig. 6; $r^2 = 0.002$, $F = 3.3$, $\beta_1 = 0.14$, $P = 0.070$). Moreover, this also abolishes the relationship between gene content and chromosomal position in this species (figs. 6 and 7; supplementary figs. 14, 15, 19, and 20, Supplementary Material online; $r^2 = -0.00090$, $F = 0.89$, $\beta_1 = -0.90$, $P = 0.35$). In tandem with patterns of repeat content, this is suggestive of a radical remodeling of genomic organization along the *C. inopinata* lineage.

Transposable Elements Are Younger in *C. inopinata* Compared with Its Close Relatives

As only a handful of transposable element superfamilies appear to underlie the atypical repetitive genomic landscape in *C. inopinata*, we suspected their evolutionary histories may likewise be exceptional. In particular, it is thought that equilibrium transposable element loads may be rare and that their evolutionary dynamics are dominated by waves of transposon proliferation and contraction corresponding to the evolution of new elements and their subsequent control by the host (Lynch 2007). If these atypical superfamilies in *C. inopinata* have emerged recently, there may not have been enough time for the host to mount an effective response, leading to their domination of the genome. To explore this possibility, we extracted Kimura distances of all transposable elements from their consensus sequences from RepeatMasker output files, and these were used as a proxy for transposable element age (as has been done in previous studies; Kapusta et al. 2017; Petersen et al. 2019;

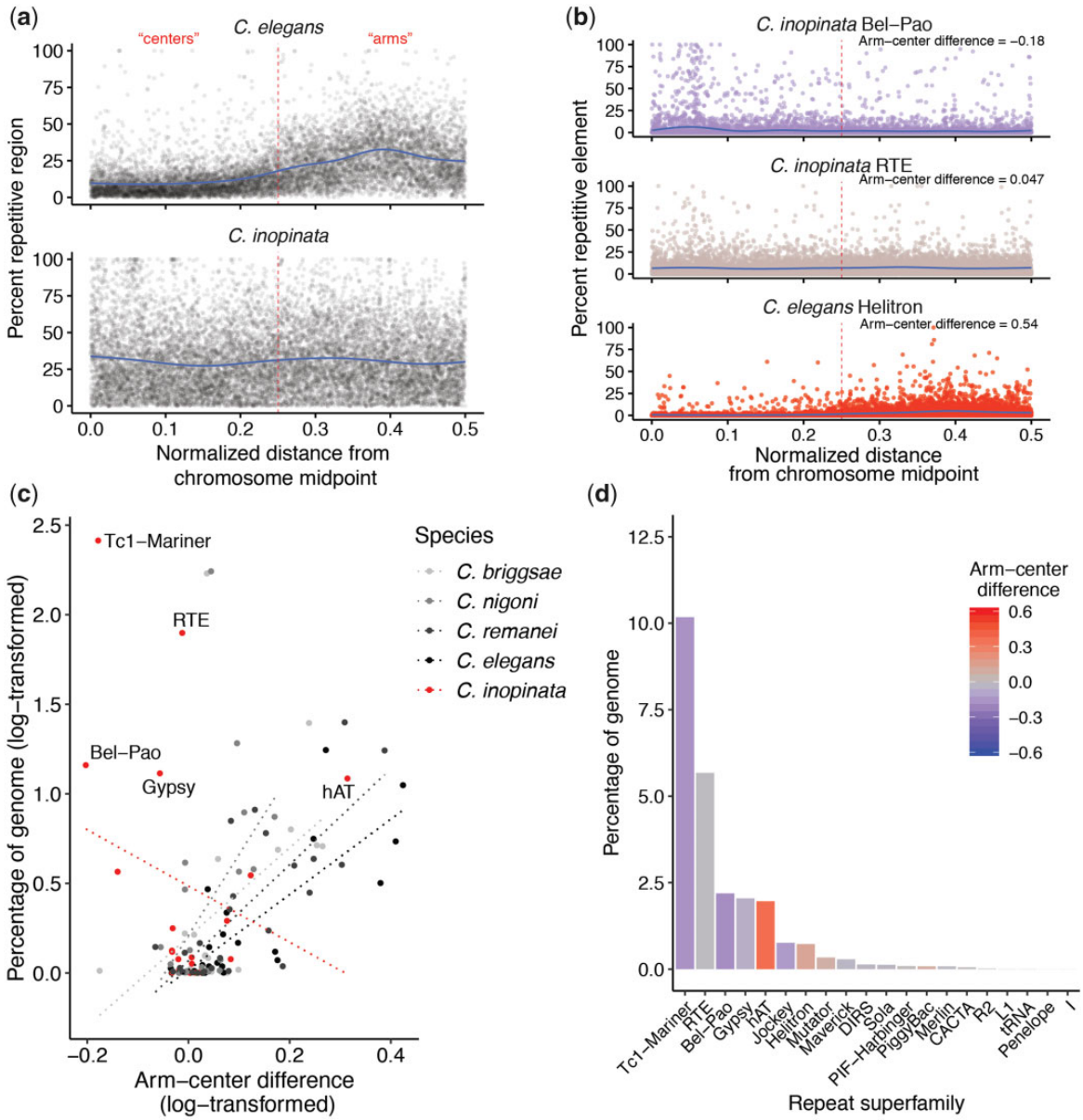

Fig. 4. The genomic structure of repeat superfamilies in *Caenorhabditis*. (*a*) The genomic landscape of repetitive elements in *Caenorhabditis elegans* and *C. inopinata* when normalized by chromosome position. Here, all genomic windows from all chromosomes are plotted, with the percentages of 10-kb windows that contain repetitive regions on the *y*-axis. "0" represents chromosome midpoints, and "0.5" represents chromosome ends. Windows can then be binned into chromosome "centers" (normalized chromosomal position <0.25 [dotted red vertical line]) or "arms" (normalized chromosomal position ≥ 0.25) to quantify the impact of chromosome position on repeat density. The blue lines were fit by generalized additive models. (*b*) Quantifying the chromosomal structure of repeat taxa. The genomic distribution of three repeat superfamilies in among two species are plotted with normalized chromosomal positions as in (*a*) (from top to bottom: Bel-Pao in *C. inopinata*, RTE in *C. inopinata*, and Helitron in *C. elegans*). Points are colored by the Cohen's *d* effect size of chromosome position (chromosome arms—centers) on repeat density with the same color gradient as in panel (*d*). Here and in all panels, this is referred to as the "arm-center difference." An effect size of one notes that the average repeat density among windows in chromosome arms is one pooled SD higher than those in centers; an effect size of zero reveals on average no difference in repeat density between chromosome arms and centers. Negative values reveal repeat densities higher in chromosome centers compared with arms. (*c*) The relationship between repeat chromosomal structure and total genomic repeat content among repeat superfamilies in five *Caenorhabditis* species. The arm-center difference is the Cohen's *d* effect size of chromosome position on repeat density as described in (*b*). The five most abundant repeat superfamilies in *C. inopinata* are labeled. All variables are log-transformed (ln(variable+1)). All linear relationships $P<0.05$ except for *C. inopinata* ($P=0.29$); additional regression statistics can be found in the text and in the Supplementary Material. Supplementary figure 13, Supplementary Material online, shows the same data but not transformed. (*d*) Repeat superfamily content in *C. inopinata*. Bars are colored by arm-center chromosome position effect size as described in panel (*b*).

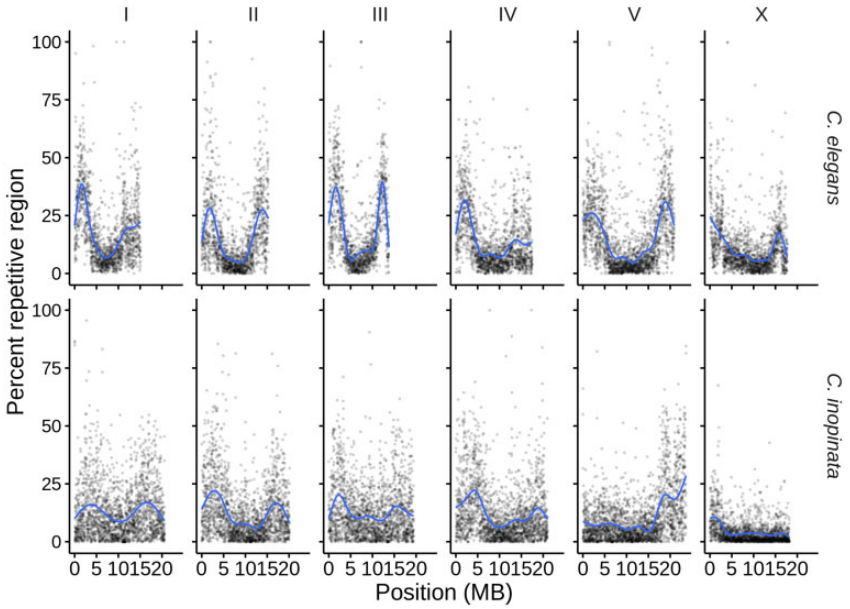

Fig. 5. *Caenorhabditis inopinata* reveals a more conventional repetitive genomic landscape when four repeat superfamilies are removed. Plotted are the percentages of 10-kb windows that contain repetitive regions by genomic position after removing Tc1-Mariner, RTE, Bel-Pao, and Gypsy repeat superfamilies. The blue lines were fit with generalized additive models.

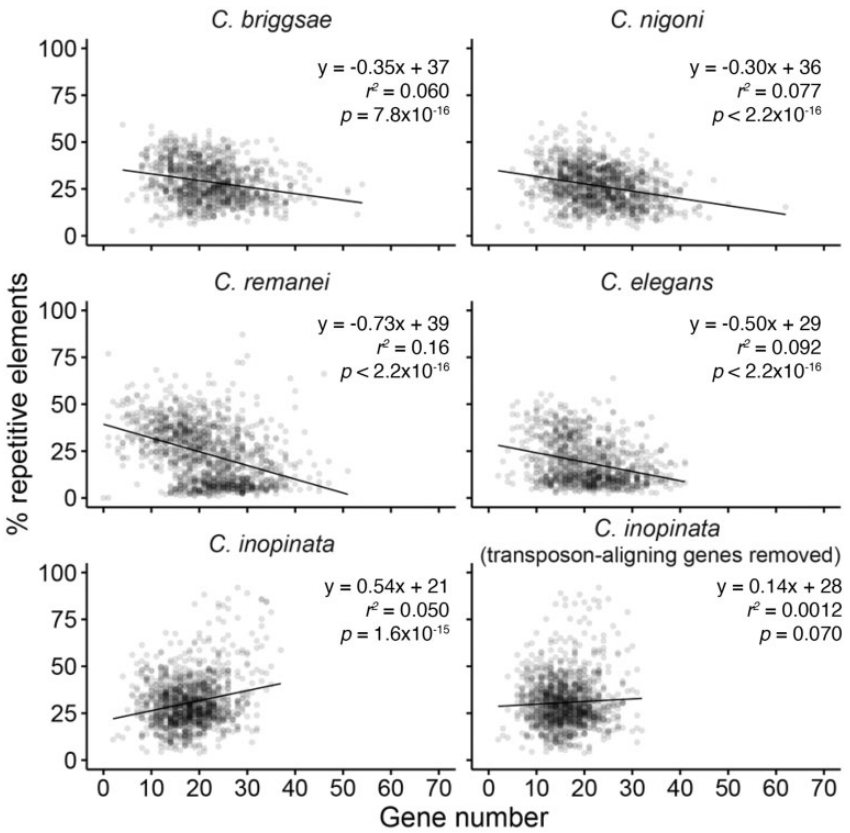

Fig. 6. The relationship between gene and repeat density in *Caenorhabditis*. Plotted are the percent repetitive region by gene count in 100-kb windows across all species. In the case of *Caenorhabditis inopinata*, an additional plot excludes 2,489 transposon-aligning genes (that also do not align to any *C. elegans* proteins) from gene counts.

Schemberger et al. 2019; see Materials and Methods). We then visualized and quantified the genomic landscapes and distributions of transposable element ages for all repeat taxa.

Globally, the genomic landscapes of transposable element age are more uniform than transposable element density for all *Caenorhabditis* species (supplementary fig. 21,

Supplementary Material online). However, differences between chromosome arms and centers are apparent in four species, although their directions are idiosyncratic. In *C. briggsae*, *C. nigoni*, and *C. remanei*, transposable elements are slightly older in chromosome arms than in chromosome centers (supplementary figs. 21 and 22, Supplementary Material online; arm-center Cohen's *d* effect sizes = 0.28, 0.12, and 0.35, respectively; linear model $P < 2.2 \times 10^{-16}$ for *C. briggsae* and *C. remanei*, $P = 1.2 \times 10^{-8}$ for *C. nigoni*). Conversely, in *C. elegans*, transposable elements are older in chromosome centers than in arms (supplementary figs. 21 and 22, Supplementary Material online; arm-center Cohen's *d* effect size = −0.21, linear model $P < 2.2 \times 10^{-16}$). In *C. inopinata*, there is no significant difference in transposable element age between chromosome arms and centers (supplementary figs. 21 and 22, Supplementary Material online; arm-center Cohen's *d* effect size = −0.0027, linear model $P = 0.27$). At the same time, the variance in transposable element age is greater in chromosome centers than in chromosome arms in all species (supplementary fig. 21, Supplementary Material online). In *C. elegans* and *C. remanei*, the variance is about two times greater in centers than in arms (*C. elegans*: $\text{variance}_{\text{centers}} = 36.4$, $\text{variance}_{\text{arms}} = 20.8$; *C. remanei*: $\text{variance}_{\text{centers}} = 42.7$, $\text{variance}_{\text{arms}} = 20.0$), whereas the difference is small or negligible in the other species (*C. briggsae*: $\text{variance}_{\text{centers}} = 27.6$, $\text{variance}_{\text{arms}} = 21.7$; *C. nigoni*: $\text{variance}_{\text{centers}} = 22.4$, $\text{variance}_{\text{arms}} = 21.0$; *C. inopinata*: $\text{variance}_{\text{centers}} = 27.0$, $\text{variance}_{\text{arms}} = 20.5$).

In general, transposable elements are younger in *C. inopinata* than in all other species (supplementary fig. 23, Supplementary Material online; *C. inopinata* mean Kimura distance = 13.6; other species = 17.0–21.2), and this pattern holds among all repeat classes and most repeat superfamilies (see supplementary_figures_index.docx, Supplementary Material online, for links to figures showing comparisons among all repeat taxa). Within *C. inopinata*, the four atypically distributed repeat superfamilies were comparable in age with taxa enriched on chromosome arms (supplementary fig. 24, Supplementary Material online). As recent waves of transposable element proliferation would be predicted to produce multimodal age distributions, this was also addressed. The only repeat superfamily in *C. inopinata* with evidence of a multimodal age distribution (Mutator; Silverman test $P = 0.02$, $k = 2$; supplementary fig. 24, Supplementary Material online) also had a conventional repetitive genomic landscape (supplementary figs. 10 and 11, Supplementary Material online). Furthermore, a recent transposable element expansion should lead to a negative relationship between transposable element insertions and transposable element age, consistent with a burst of young transposable element activity. This also was not observed within *C. inopinata* transposable element superfamilies (supplementary figs. 25–27, Supplementary Material online). Taken together, although repetitive elements are younger in *C. inopinata* compared with its close relatives (supplementary fig. 23, Supplementary Material online), this suggests that the repeat superfamilies underlying *C. inopinata*'s atypical repetitive genomic landscape were established in the distant past after this species' divergence from *C. elegans*.

Simulations Reveal Chromosomal Heterogeneity in Insertion Fitness Effects and Recombination Can Promote Genomic Repetitive Landscapes

Recombination rate also covaries with chromosomal position in nematodes and is elevated in chromosome arms relative to centers in both *C. elegans* and *C. briggsae* (Rockman and Kruglyak 2009; Ross et al. 2011). How does intrachromosomal variation in recombination rate contribute to genomic repetitive organization? To address this question, we implemented forward simulations of transposable element evolution with the SLiM software package (Haller and Messer 2019; fig. 8; see Materials and Methods). Primarily, chromosomes with either a uniformly high recombination rate or with three domains of differing recombination rate (high in chromosome arms and low in centers) were modeled. In addition, various patterns of chromosomal heterogeneity in the fitness effects of transposable element insertion were also addressed. Intrachromosomal variation in recombination can cause transposable elements to be enriched on higher-recombining chromosome arms depending on insertion fitness effects (fig. 8*a*–*c*). When transposable elements are neutral (fig. 8*a*) or highly deleterious (fig. 8*c*) across the genome, uniform repetitive genomic landscapes emerge (arm-center Cohen's *d* effect size = −0.038 to 1.9×10^{-4}; Wilcoxon rank sum test $W = 27678635–282227801$, $P = 0.093–0.69$). However, when transposable elements are weakly deleterious, repetitive elements are enriched in chromosome arms when domains of recombination are present (arm-center Cohen's *d* effect size = 1.1; Wilcoxon rank sum test $W = 44504020$, $P < 2.2 \times 10^{-16}$) but are uniformly distributed when there is uniform recombination (fig. 8*b*; arm-center Cohen's *d* effect size = 0.024; Wilcoxon rank sum test $W = 28545691$, $P = 0.11$). This pattern is also observed (with a weaker effect) when beneficial mutations are introduced (fig. 8*g*). As in these simulations insertion fitness effects were drawn from a gamma distribution (see Materials and Methods), we also simulated populations with fixed insertion effects and three chromosomal domains of differing recombination rate (supplementary fig. 28, Supplementary Material online). Here, only weak insertion fitness effects ($s = -0.0002$) are sufficient to generate genomic landscapes where transposable elements are enriched on chromosome arms (supplementary figs. 28 and 29, Supplementary Material online; arm-center Cohen's *d* effect size = 0.42; Wilcoxon rank sum test $W = 34789102$, $P < 2.2 \times 10^{-16}$). With more deleterious insertions, transposable elements become enriched on lower-recombining chromosome centers (supplementary figs. 24 and 25, Supplementary Material online; arm-center Cohen's *d* effect size = −0.51 to −0.48; Wilcoxon rank sum test $W = 19785003–21767588$, $P < 2.2 \times 10^{-16}$ for all). Thus, weakly deleterious transposable elements can interact with intrachromosomal variation in recombination rate to drive the evolution of repetitive genomic organization.

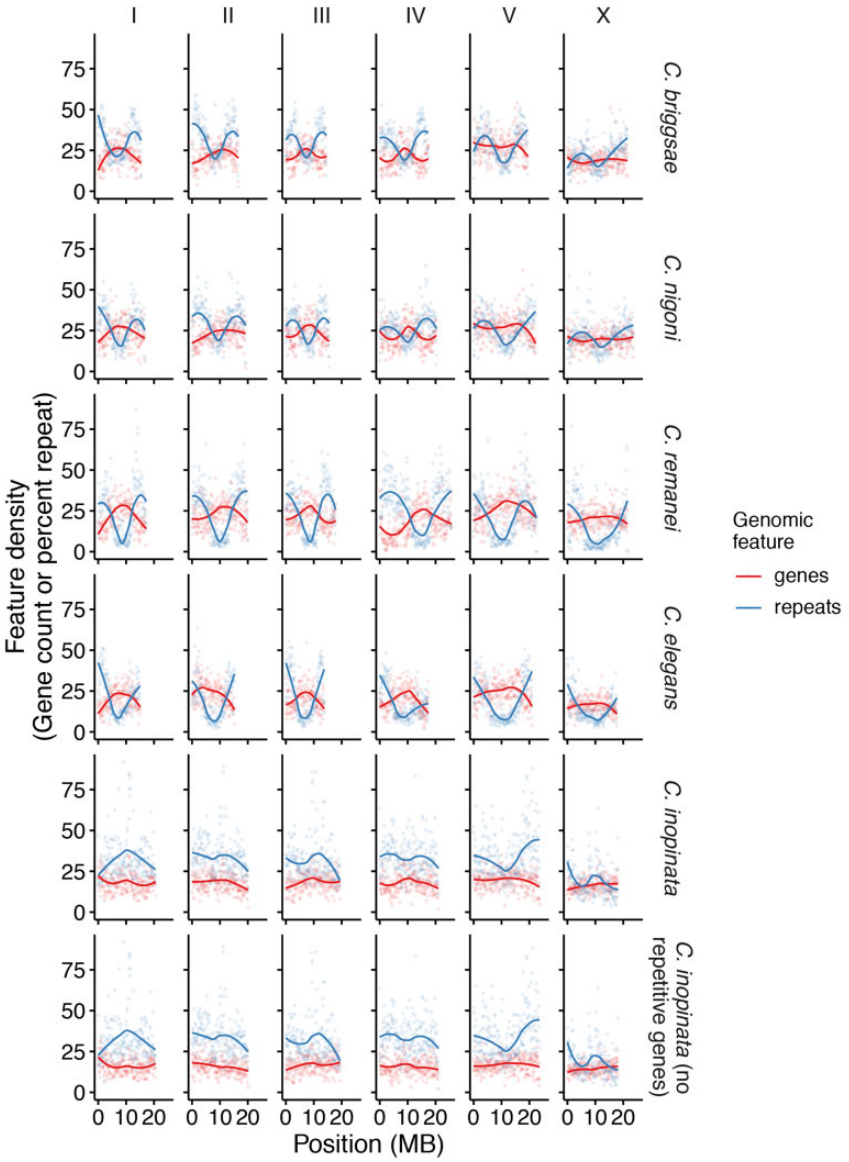

Fig. 7. Genomic landscapes of gene and repeat density in *Caenorhabditis*. Columns represent the six chromosomes; rows represent species. Here, the gene densities of *Caenorhabditis inopinata* when 2,489 transposon-aligning genes are excluded are also plotted (last row; the repeat densities are identical to the row above); 100-kb windows are plotted. For repeats, densities are reported as the percentage of the window that is repetitive. For genes, densities are reported as the number per window because gene length also covaries with genomic position (supplementary fig. 35 and 36, Supplementary Material online). Lines were fit by LOESS local regression.

However, protein-coding genes are also enriched in chromosome centers in most *Caenorhabditis* species (fig. 7), which suggests that there may also be chromosomal heterogeneity in the fitness effects of transposable element insertions. Specifically, because a transposable element is more likely to abrogate gene function in gene-rich chromosome centers, it may have a greater fitness consequence upon insertion. We then also simulated populations where the fitness effect of transposable element insertion is greater in chromosome centers than arms (fig. 8*d–f*). Notably, heterogeneity in insertion fitness effects consistently revealed repetitive genomic landscapes with transposable elements enriched in chromosome arms (fig. 8*d–f*; arm-center Cohen's *d* effect size = 0.24–2.2; Wilcoxon rank sum test $W = 31994891–50127409$, $P < 2.2 \times 10^{-16}$ for all). These patterns emerged even in the absence of heterogeneity in recombination rate variation (fig. 8*d–f*). This pattern holds for both "copy-and-paste" as well as "cut-and-paste" (fig. 8*f*) models of transposable element replication. Additionally, as *Caenorhabditis* exhibits variation in reproductive mode with both female/male and selfing species, we also simulated selfing populations with chromosomal heterogeneity in insertion fitness effects. These also revealed repetitive genomic landscapes with transposons enriched in chromosome arms, regardless of recombination rate variation (supplementary fig. 30, Supplementary Material online; arm-center Cohen's *d* effect size = 1.3 for both simulations; Wilcoxon rank sum test $W = 46248175$, 46162860, $P < 2.2 \times 10^{-16}$ for both). Such landscapes were not observed in selfing simulations with uniformly weak

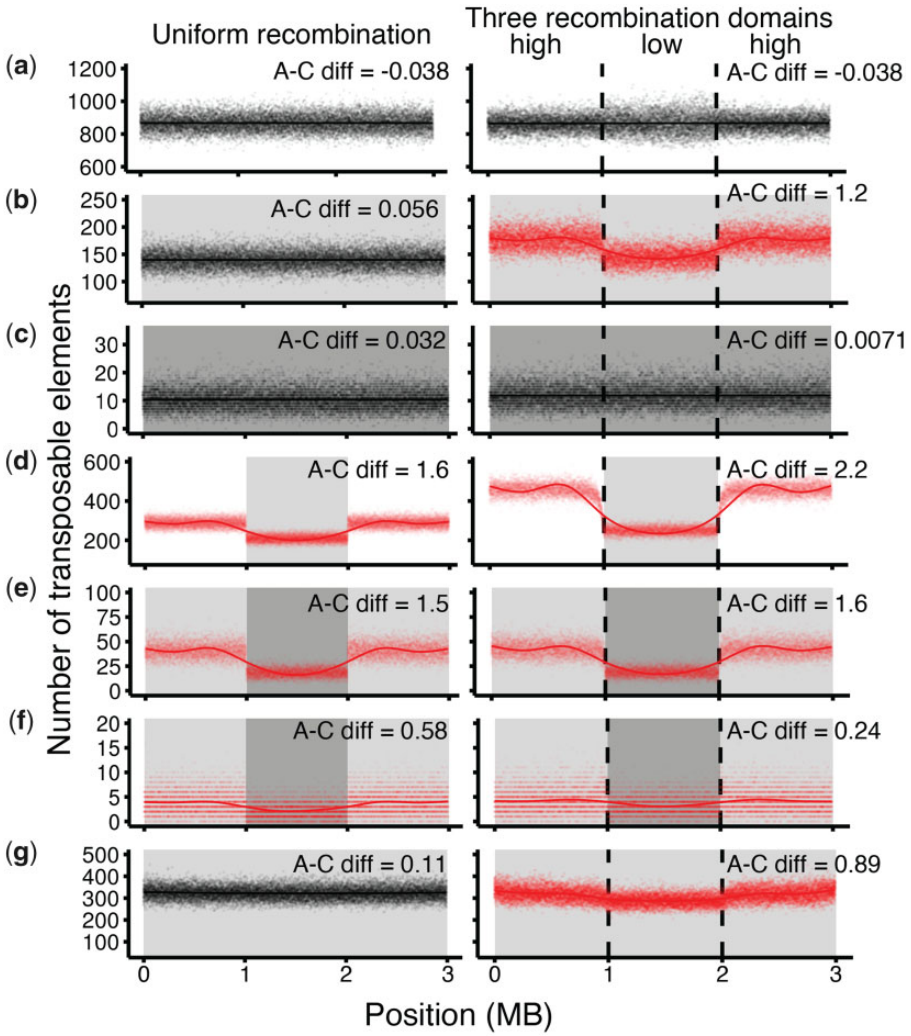

Fig. 8. Simulations reveal chromosomal heterogeneity in insertion fitness effects is sufficient for promoting genomic repetitive organization. 3 MB chromosomes were evolved under multiple evolutionary scenarios. Plotted are the total number of transposable element sites in 10-kb non-overlapping genomic windows along the chromosome after 50,000 generations. Each point represents the total number of transposable element sites in a population in that window. The left column shows scenarios where there were uniformly high rates of recombination along the chromosome ($r = 5 \times 10^{-7}$); the right column shows scenarios where there were three chromosomal domains of recombination (boundaries denoted by the dashed lines; chromosome arms $r = 5 \times 10^{-7}$; chromosome centers $r = 1 \times 10^{-9}$). Gray boxes represent chromosomal regions with deleterious fitness effects of transposable element insertion (light gray, mean $s = -0.0006$; dark gray, mean $s = -0.03$; white, $s = 0$ for all transposable element insertions). When not neutral, fitness effects of insertion were drawn from a gamma distribution. Thick lines were fit by LOESS regression. Red lines represent scenarios with arm-center Cohen's *d* effect sizes >0.2; black lines show scenarios with no or negligible chromosomal organization of repeats (arm-center Cohen's *d* effect sizes <0.2). "A-C diff" is the arm-center Cohen's *d* effect size of transposable element number between chromosome arms and centers. All scenarios implemented copy-and-paste models of transposable element replication with the exception of (*g*), which used a cut-and-paste model. Brief descriptions of each scenario are described as follows (see Materials and Methods for details): (*a*) All transposable element insertions neutral (these populations are not at equilibrium, supplementary fig. 37, Supplementary Material online), (*b*) All transposable element insertions weakly deleterious. (*c*) All transposable element insertions highly deleterious. (*d*) Transposable element insertions weakly deleterious in center; transposable element insertions neutral in arms. (*e*) Transposable element insertions highly deleterious in center; transposable element insertions weakly deleterious in arms. (*f*) Transposable element insertions highly deleterious in center; transposable element insertions weakly deleterious in arms; cut-and-paste mode of TE replication. (*g*) All transposable element insertions weakly deleterious; beneficial mutations introduced ($\mu = 1 \times 10^{-9}$).

selection along the chromosome (supplementary fig. 31, Supplementary Material online; arm-center Cohen's *d* effect size = −0.0075, −0.025; Wilcoxon rank sum test $W = 28108465$, 27641416, $P = 0.95$, 0.069). Then, whereas chromosomal heterogeneity in recombination rate alone can generate repetitive genomic landscapes when insertions are weakly deleterious, chromosomal heterogeneity in insertion fitness effects can promote such landscapes across a variety of evolutionary scenarios.

Transposable element ages were also recorded in these simulations. In *Caenorhabditis*, chromosome centers have higher variance in transposable element age than chromosome arms (supplementary fig. 32, Supplementary Material online). Furthermore, there is variation in the direction and

extent of arm-center differences in transposable element age among *Caenorhabditis* species (supplementary figs. 21 and 22, Supplementary Material online). Likewise, there is idiosyncrasy among the genomic landscape of transposable element ages in simulations (supplementary fig. 32–34, Supplementary Material online). Notably, both chromosomal heterogeneity in recombination rate and insertion fitness effects can lead to increased age variance in chromosome centers (supplementary fig. 32, Supplementary Material online), and most scenarios with transposable elements enriched in chromosome arms had more variance in insertion age in chromosome centers (supplementary fig. 32, Supplementary Material online). Thus the various processes that can lead to repetitive genomic landscapes often promote the increased variance of insertion age in chromosome centers relative to arms.

Discussion

Mobile elements are known to harbor tremendous diversity in their replication strategies. However, it remains unclear to what extent repeat taxon-specific evolution (see supplementary_tables.xlsx, sheet 23 for repeat taxonomy, Supplementary Material online) shapes the abundance and chromosomal distribution of global repetitive content among genomes. Here, we described just these patterns of transposable element content in five *Caenorhabditis* assemblies, and repeat taxon-specific evolution among lineages appears widespread in this group. Diversity in the abundances of various repeat superfamilies is common. For instance, the Jockey superfamily has expanded in the *C. remanei* lineage; the Helitron superfamily of DNA transposons is uniquely abundant in the *C. briggsae*/*C. nigoni* clade (supplementary fig. 8, Supplementary Material online). There is broad variation in repeat content both among and within genomes across the range of transposable element taxa (fig. 3; supplementary figs. 6–12, Supplementary Material online). Such chromosomal heterogeneity in repeat density depending on the type of transposable element has also been observed in multiple systems, including humans (Myers et al. 2005), mice (Shifman et al. 2006), and maize (Stitzer et al. 2019). However, here elements generally tend to be enriched on chromosome arms relative to centers as has observed in numerous studies in *Caenorhabditis* (figs. 2 and 4*c*; supplementary fig. 13, Supplementary Material online; *C. elegans* Sequencing Consortium 1998; Stein et al. 2003; Fierst et al. 2015; Yin et al. 2018).

Caenorhabditis inopinata obviously bucks this general trend in repeat organization. Although its high repeat content has been previously noted (Kanzaki et al. 2018), here we describe its exceptional repetitive landscape (this previous work did describe *C. inopinata*'s genomic landscape of tandem repeats, which follows a typical chromosomal distribution that we also observed [here described as simple repeats, supplementary figs. 6 and 7, Supplementary Material online]). Despite this divergent genomic organization, some repeat taxa in this species exhibit the conventional enrichment on chromosome arms, primarily hAT, Mutator, and Helitron DNA transposons (fig. 4*d*; supplementary fig. 10, Supplementary Material online). Additionally, repetitive regions not obviously related to transposable elements (such as Satellite DNA, low-complexity repeats, and unclassified repetitive elements) are also enriched on chromosome arms in *C. inopinata* (supplementary fig. 6, Supplementary Material online). Thus, the breakdown in the repetitive genomic landscape in this species is repeat taxon-specific. The exclusion of four repeat superfamilies that have high abundance and atypical genomic organization reveal a more conserved repeat distribution (figs. 4*c*, 4*d*, and 5). Notably, these particular repeat superfamilies vary greatly in their replicative diversity. Tc1/Mariner elements are cut-and-paste class II DNA transposons (Wicker et al. 2007). RTE is a class I LINE retrotransposon that uses its own transcription product to prime reverse transcription (Wicker et al. 2007). Bel-Pao and Gypsy are more complex LTR retrotransposons whose autonomous copies usually encode retroviral proteins and use tRNA's to prime reverse transcription (Wicker et al. 2007). Thus, variation in the abundance and chromosomal distribution of diverse repeat taxa underlies change in repetitive landscapes.

The repetitive genomic landscape of *C. inopinata* is so exceptional that the possibility it arises from assembly errors must be considered. Conventional measures of assembly quality (N50; number of scaffolds; number and length of gaps; and BUSCO completeness score, among others) all suggest assembly errors are not driving these patterns as this is a highly contiguous and almost entirely complete assembly (supplementary fig. 1 and supplementary_tables. xls sheet 1, Supplementary Material online). Indeed, the *C. briggsae* assembly is more fragmented than the *C. inopinata* assembly yet maintains a conserved repetitive genomic landscape (supplementary fig. 1, Supplementary Material online and fig. 2). Moreover, that *C. inopinata* reveals a more conventional repetitive genomic landscape when four repeat superfamilies are removed also suggests that these patterns are biologically relevant (fig. 5; also see the repeat count genomic landscape, supplementary_figures_index.docx, Supplementary Material online). Furthermore, the *C. inopinata* assembly reveals patterns of chromosomal synteny consistent with the other *Caenorhabditis* assemblies when considering single-copy orthologous genes (supplementary fig. 38, Supplementary Material online). All pairwise species comparisons (including those with *C. inopinata*) reveal $\geq$96% of single-copy orthologs to be on the same chromosome in both species (supplementary fig. 38, Supplementary Material online). However, most pairwise species comparisons also reveal widespread inversions and translocations within chromosomes (with the notable exception of the *C. briggsae*–*C. nigoni* sister species; supplementary fig. 38, Supplementary Material online), consistent with previous observations (Stein et al. 2003; Fierst et al. 2015; Kanzaki et al. 2018; Yin et al. 2018; Teterina et al. 2020); their conserved repetitive genomic landscapes persist despite these rearrangements (fig. 2). *Caenorhabditis inopinata* does not have an exceptional degree of between-chromosome translocation nor within-chromosome rearrangements despite its unusual repetitive genomic landscape

(supplementary fig. 38, Supplementary Material online; fig. 2). Finally, the recently published assembly of the cattle-associated *Caenorhabditis bovis* also reveals it to have a more uniform repetitive genomic landscape, suggesting this may be a biologically reproducible pattern within this group (Stevens et al. 2020). Thus, it is highly likely that the chromosomal distributions of repetitive elements observed in *C. inopinata* are not due to errors in genome assembly.

This chromosomal heterogeneity in transposable element distribution among nematode chromosomes has been widely noted (Duret et al. 2000; Rizzon et al. 2003; Cutter et al. 2009), and these patterns covary with multiple genomic features that are also nonrandomly distributed along chromosomes (Cutter et al. 2009). In *C. elegans*, recombination rate is elevated on chromosome arms (Rockman and Kruglyak 2009), and protein-coding genes are enriched on chromosome centers (*C. elegans* Sequencing Consortium 1998). Heterochromatic regions (Garrigues et al. 2015) and piRNA's (Shi, et al. 2013) are also enriched on chromosome arms; consistent with this, chromosome arms are less transcriptionally active than chromosome centers (Garrigues et al. 2015). As heterochromatic marks and small RNA's have been connected to transposon regulation in other systems (Lee 2015), transposable elements may proliferate in such genomic regions where insertions are less costly and easier to control. Further, essential genes are enriched in chromosome centers in *C. elegans* (Kamath et al. 2003). And, chromosome arms are associated with each other in physical space and tend to be tethered to nuclear membranes (Cabianca et al. 2019). Thus, various genomic features covary with one another, but the outstanding question remains: what is cause and what is effect with respect to the distribution of these various genomic features? Such patterns of these genomic features have been observed across *Caenorhabditis* (Stein et al. 2003; Ross et al. 2011; Fierst et al. 2015; Kanzaki et al. 2018; Yin et al. 2018). Taken together, this suggests that this chromosomal organization is an ancient, defining characteristic of the *Caenorhabditis* genome. This may also explain the counter-intuitive observation that repetitive elements are enriched in high recombining regions in this system, whereas opposite patterns are observed in other systems (Rizzon et al. 2002; Wright et al. 2003; Jensen-Seaman et al. 2004; Liu et al. 2009; Pan et al. 2011) and are predicted by evolutionary theory (Hill and Robertson 1966; Langley et al. 1988). Although it is difficult to disentangle cause and effect, the heterochromatic, less gene-dense chromosomal regions have higher rates of recombination; repetitive elements accumulate there perhaps because of their potentially lower fitness effects upon insertion.

We performed evolutionary simulations to understand how heterogeneity in recombination rate and insertion fitness effects along the chromosome can influence repetitive genomic landscapes (fig. 8). We found that chromosomal heterogeneity in recombination rate alone is sufficient to generate the enrichment of transposable elements in high-recombining chromosome arms only when such elements have weakly deleterious fitness effects upon insertion (fig. 8*b*). When insertions are uniformly neutral (fig. 8*i*) or highly deleterious (fig. 8*c*), such patterns are not observed or insertions become enriched on *low*-recombining regions (supplementary figs. 28 and 29, Supplementary Material online). This is largely consistent with previous evolutionary simulations that suggested that extremely low recombination rates are needed for Hill–Robertson effects to fix transposable elements (Dolgin and Charlesworth 2008). Conversely, chromosomal domains of differing fitness effects were sufficient to promote such patterns in a range of evolutionary scenarios (fig. 8*d–f*), including those with a selfing mode of reproduction (supplementary fig. 30, Supplementary Material online). Notably, nematode genomic landscapes of repetitive element age reveal increased variance in chromosome centers (supplementary fig. 21, Supplementary Material online), which is comparable with patterns of transposable element age in simulations that also have structured repetitive genomic landscapes (supplementary fig. 32, Supplementary Material online). This suggests that the genomic landscapes of repetitive element abundance and repetitive element age that we observe in nematodes may be caused by the same evolutionary processes.

As chromosomal heterogeneity in transposable element fitness effects can robustly generate repetitive genomic landscapes, this suggests that stronger purifying selection on transposable element insertions in the genomic clusters of genes in the centers of chromosomes may be driving the conserved repetitive genomic landscapes in *Caenorhabditis*. Indeed, *C. inopinata* does not have protein-coding genes enriched in the centers of chromosomes like its close relatives (fig. 7; supplementary figs. 16–20, Supplementary Material online), which is consistent with its divergent repetitive genomic landscape and this hypothesis. However, estimates of repetitive element ages provide no evidence that such superfamilies are younger than repeat taxa that have more conventional genomic landscapes (supplementary figs. 24–27, Supplementary Material online). But when all repeat content is taken into account (aside from simple repeats and satellite sequences), repetitive elements are younger in *C. inopinata* compared with its close relatives (supplementary figs. 21–23, Supplementary Material online). This suggests that after its divergence from *C. elegans*, the *C. inopinata* lineage may have experienced a proliferation of repetitive elements across multiple repeat superfamilies at around the same time. However, this assumes that pairwise sequence distances are comparable proxies for element age across repetitive element types and species, and different models of sequence evolution for such variant types are likely needed for more precise and accurate estimates of element age.

It remains unclear exactly why certain abundant repeat superfamilies have atypical chromosome distributions despite being comparable in age with those that are enriched in chromosome arms in *C. inopinata*. The role of small RNA regulators on the evolution of *Caenorhabditis* repetitive landscapes, and these types of elements in particular, is also an important open question because a number of such factors have apparently been lost in *C. inopinata* (Kanzaki et al. 2018). Particularly, three genes in the ERGO-1 siRNA pathway (*ergo-1*, *eri-6/eri-7*, and *eri-9*; Piatek and Werner 2014) are not present in this species (Kanzaki et al. 2018). As *ergo-1* encodes an

RNA-silencing argonaut protein (Yigit et al. 2006), and as small RNA's are known to regulate transposable elements (Piatek and Werner 2014; Fischer and Ruvkun 2019), the loss of these genes in *C. inopinata* poses a tantalizing line of future research for understanding its repetitive landscape. Furthermore, the demographic and evolutionary history of a lineage can drive variation in the distributions of transposable elements (Lockton et al. 2008), and this remains unexplored in *C. inopinata*. In any case, future work on recombination and the genomic organization of various features in *C. inopinata* (such as heterochromatin, spatial genomic structure, and transcription, among others) within their evolutionary context will be needed to better understand the causes of its divergent repetitive genomic landscape.

C. inopinata was not a lone outlier in one respect—*C. japonica* was also observed to have a high percentage of its gene set aligning to transposable elements (12%; supplementary fig. 18, Supplementary Material online). Previously observed to have high transposable element content (Fierst et al. 2015; Szitenberg et al. 2016), *C. japonica* shares notable ecological features with *C. inopinata*. They have only been observed on east Asian islands—*C. japonica* on Kyushu Island in Japan (Kiontke et al. 2002; Yoshiga et al. 2013) and *C. inopinata* in Okinawa and Taiwan (Kanzaki et al. 2018; Woodruff and Phillips 2018). Moreover, they are both phoretic host specialists. *Caenorhabditis* nematodes generally thrive on rotting plant material (Kiontke et al. 2011), and they travel from one resource patch to another on invertebrate carriers (Schulenburg and Félix 2017). Many *Caenorhabditis* species are promiscuous with respect to their hosts (including *C. elegans*, *C. briggsae*, *C. remanei*; Cutter 2015), whereas others appear to be host specialists, having only been found with one invertebrate carrier. *Caenorhabditis japonica* is associated with the Japanese shield bug *Parastrachia japonensis* (Okumura et al. 2013; Yoshiga et al. 2013), and *C. inopinata* disperses on *Ceratosolen* fig wasps (Kanzaki et al. 2018; Woodruff and Phillips 2018). The proliferation of transposable elements in this group may then somehow connected to host specialization. However, some presumptive host specialists do not have obvious expansion of transposable element content among their predicted protein-coding genes (*C. angaria*, *C. plicata*, and *C. monodelphis*; supplementary fig. 18, Supplementary Material online; Volk 1951; Sudhaus et al. 2011; Slos et al. 2017), and the extent of host preferences among *Caenorhabditis* species remains largely unknown (Cutter 2015). Additionally, there are many potentially island-restricted *Caenorhabditis* species whose genomes have yet to be assembled (Cutter 2015). Thus, the generation of more genome assemblies, in tandem with further exploration of *Caenorhabditis* ecology, will be needed to frame this question in its proper comparative phylogenetic context and address the possible role of ecological specialization in the evolution of transposable elements.

Such detailed phylogenetic comparative methods using more taxa would also be needed to address the obvious question regarding the impact of reproductive mode on transposable element content (Glémin et al. 2019). Evolutionary theory predicts conflicting expectations regarding the abundance of transposable elements in selfing lineages (Wright and Schoen 1999; Morgan 2001), with different models of selection leading to either transposon expansion or reduction. Additionally, selfing lineages of *Arabidopsis* harbor more transposable elements than outcrossers (Wright, et al. 2001; Lockton and Gaut 2010). But in agreement with previous results in *Caenorhabditis* (Fierst et al. 2015; Yin et al. 2018), we find no obvious pattern among transposable element content and reproductive mode, although our phylogenetic sample is quite small. This is also consistent with observations in asexual arthropods (Bast et al. 2016) and bdelloid rotifers (Nowell et al. 2018) that reveal no amplification of repeat content upon change in reproductive mode. As the *Caenorhabditis* genus only has three independent transitions to self-fertile hermaphroditism (Stevens et al. 2019), this group may not be well suited for addressing this question in a comparative phylogenetic context. However, a study interrogating the relationship between reproductive mode and transposable elements content in nematodes among a broader phylogenetic sample found no clear predictors of transposable element abundance and concluded genetic drift, independent of reproductive mode, is a major driver of transposon evolution in nematodes (Szitenberg et al. 2016). Although we also see no clear association of transposable element abundance with predicted population sizes here, as they range across at least two orders of magnitude in the *Caenorhabditis* genus (Dey et al. 2013; Fierst et al. 2015), future comparative phylogenetic approaches can also be used to test the prediction that genetic drift should influence transposable element abundance.

Regardless of the forces underlying their proliferation, the exceptional transposable element expansion in the morphologically exceptional *C. inopinata* lineage cannot be ignored. Whereas many *Caenorhabditis* species are largely morphologically indistinguishable, *C. inopinata* is twice as long and develops twice as slowly as its close relatives (Kanzaki et al. 2018; Woodruff et al. 2018, 2019), and it thrives in the lumen of fresh figs instead of on rotting plant material (Kanzaki et al. 2018; Woodruff and Phillips 2018). Its divergent genomic organization is likewise striking. Repetitive elements have been shown to be a source of adaptive mutations in numerous contexts—peppered moths (van't Hof et al. 2016), oranges (Butelli et al. 2012), and stickleback fish (Ishikawa et al. 2019) all have evidence of such beneficial insertions. In addition to laying the groundwork for understanding the origins, maintenance, and stability of genome structure in the face of rampant mobile element proliferation, this work also sets the stage for understanding how mobile elements can promote or constrain rapid morphological and ecological change.

Materials and Methods

Genome Assemblies

Five *Caenorhabditis* genome assemblies with chromosome-level contiguity were used for this study (fig. 1 and supplementary fig. 1 and supplementary_tables.xls sheet 1 for

assembly statistics, Supplementary Material online). The *C. elegans*, *C. briggsae*, and *C. nigoni* (Yin et al. 2018) genomes were retrieved from WormBase ParaSite (parasite.wormbase.org; last accessed May 5, 2020; Howe et al. 2017). The *C. inopinata* genome (Kanzaki et al. 2018) was retrieved from the *Caenorhabditis* Genomes Project (caenorhabditis.org; last accessed May 5, 2020; Slos et al. 2017; Stevens et al. 2019). A new chromosome-level assembly of the *C. remanei* genome was also used (this new assembly can be found at www.ncbi.nlm.nih.gov/assembly/GCA_010183535.1; last accessed May 5, 2020; Teterina et al. 2020). Protein sets of 28 *Caenorhabditis* species and *Diploscapter coronatus* were retrieved from the *Caenorhabditis* genomes project. Versions of all genome assemblies and protein sets can be found with the deposited data and code associated with this work (https://github.com/gcwoodruff/transposable_elements_2019; last accessed May 5, 2020).

Repeat Inference and Quantification

After excluding mitochondrial DNA, assemblies were masked and annotated for repeat content through a hybrid approach using multiple software packages (inspired by Coghlan et al. 2018 and avrilomics.blogspot.com/2015/09/lrtharvest.html; last accessed May 7, 2020; supplementary fig. 2, Supplementary Material online). De novo repeat libraries were generated with RepeatModeler (options: -engine ncbi -pa 16; Benson 1999; Bao and Eddy 2002; Price et al. 2005). Concurrently, sequences that align to transposable elements were detected with TransposonPSI using default parameters (transposonpsi.sourceforge.net; last accessed May 7, 2020). LTRHarvest was also used to identify LTR retrotransposon sequences in these assemblies (options: -seqids yes -tabout no -gff3; Ellinghaus et al. 2008). LTRHarvest output was further filtered (LTRHarvest option: -hmms) to extract just those sequences containing only LTR retrotransposon domains; this was done with Pfam (Finn et al. 2016) Hidden Markov Models of known LTR retrotransposon domains reported in (Steinbiss et al. 2009). The RepeatModeler, TransposonPSI, and LTRHarvest species-specific repeat libraries were then concatenated with additional repeat libraries. One of these is the Rhabditida library ("Rhabditida.repeatmasker") associated with the RepeatMasker software (Tarailo-Graovac and Chen 2009). Additionally, the *C. elegans* and *C. briggsae* repeat libraries ("cbrapp.ref," "celapp.ref," "cbrrep.ref," and "celrep.ref") from RepBase (Bao et al. 2015) were also combined with the above libraries to generate redundant repeat libraries for each assembly. USEARCH (Edgar 2010) was then used to cluster repeats (options: -id 0.8 -centroids -uc -consout -msaout) and generate nonredundant libraries. These repetitive sequences were then classified with RepeatClassifier (part of the RepeatModeler software) with default parameters. Repeats classified as "Unknown" were then aligned to *Caenorhabditis* species-specific protein sets with BlastX (options: -num_threads 24 -outfmt 6 -evalue 0.001; Camacho et al. 2009); unclassified repeats that aligned to predicted proteins were then removed from the repeat libraries. These subsequent repeat libraries were then used with RepeatMasker (part of the RepeatModeler software; options: -s -gff -pa 16) to mask the genome assemblies. Libraries were generated independently for each species; genomes were masked with species-specific libraries (i.e., they were not combined before the final masking).

To measure and visualize the global landscape of repeats, Bedtools *nuc* was used to measure the fraction of masked bases in nonoverlapping windows across the genomes (both 10 kb and 100 kb window size; default parameters). This was also done for all specific repeat classes, orders, superfamilies, and families. Repeats classified by RepeatClassifier (i.e., repeats not classified as "Unknown") were annotated by a custom repeat taxonomy (supplementary_tables.xlsx, sheet 23, Supplementary Material online) informed by the classification system in (Wicker et al. 2007). Here, we refer to "repeat taxa" as such groups or types of a repetitive element defined by this taxonomy. Repeat classes, orders, superfamilies, and families are repeat taxonomic ranks, and "class I retrotransposons," "LTR," "Bel-Pao," and "Pao" are examples of repeat taxa within those respective taxonomic ranks. Repetitive elements classified as "low complexity," "simple repeats," and "satellite" are included here as a matter of completeness but should be interpreted with caution because of their difficulty to assemble and annotate. To measure the general trend of repeat density along all chromosomes simultaneously, windows were normalized by chromosome position using custom Linux scripts (all code associated with this work has been deposited https://github.com/gcwoodruff/transposable_elements_2019) by setting the median chromosome base pair to 0 and the end chromosome base pairs to 0.5. To measure the impact of specific repeat families on the global landscape of repetitive elements, specific families were removed from the GFF file generated by RepeatMasker with custom Linux scripts, and assemblies were remasked with Bedtools *maskfasta* with annotations excluding each specific family. Masked assemblies were then processed as above to quantify and visualize global repeat landscapes.

To understand the extent of transposon representation in protein-coding genes, all predicted protein sequences from 28 *Caenorhabditis* genomes were aligned to the TransposonPSI (transposonpsi.sourceforge.net) transposon protein database ("transposon_db.pep") using BlastP (options: -outfmt 6 -evalue 0.001; Camacho et al. 2009). Unique proteins that aligned to transposons were extracted and counted with custom Linux scripts to determine the fraction of transposon-aligning protein-coding genes per genome. In the case of *C. inopinata*, its transposon-aligning protein set was aligned to the set of *C. elegans* proteins (with its 100 transposon-aligning proteins removed) in the same manner to find the intersection of protein-coding genes that align to both transposons and to otherwise homologous nematode proteins. Gene densities and gene lengths along chromosomes were determined with genome annotation files and summed (gene counts) or averaged (gene lengths) across 10-kb and 100-kb genomic windows with bedtools *map*. Single-copy orthologs were also inferred with OrthoFinder2 (Emms and Kelly 2018); genic positions were extracted from annotations to visualize synteny patterns.

Measures of repetitive element divergence were extracted from ".align" files generated by RepeatMasker for each species. These contain alignments of all repetitive genomic elements with their consensus sequences as well as various measures of divergence for each alignment. Among these are Kimura distances (with correction for CpG pairs), which were used as a proxy for element age (as has been done in previous studies, Kapusta et al. 2017; Petersen et al. 2019). Kimura distances were not converted to strict ages due to known dating issues in *Caenorhabditis* nematodes (Cutter 2008; Cutter et al. 2019); these are likely exacerbated when using rapidly evolving transposable elements. Kimura distances for each alignment were joined to the repeat annotations generated by RepeatMasker with bedtools *intersect* (options: -wao -a). bedtools *map* (options -o mean) was used to estimate mean repeat element Kimura distances across 10-kb genomic windows for all repeat classes, orders, superfamilies and families. Genomic landscapes, arm-center effect sizes, and other summary statistics were visualized and analyzed in a manner analogous to repeat density as above. Code and data for all divergence estimates and analyses have been deposited (https://github.com/gcwoodruff/transposable_elements_2019).

Simulations of Transposable Element Evolution

Simulations of transposon element evolution were conducted in SLiM 3.3 (Haller and Messer 2019) with a script based on recipe 13.6 (Modeling transposon elements) from the SLiM 3.0 manual (February 28, 2018 revision). In the recipe, active transposons that able to "copy and paste" themselves were simulated. Briefly, base SLiM simulates Wright–Fisher populations with diploid, chromosome-bearing individuals. Chromosomes are composed of discrete positions that can accumulate mutations which can then segregate in evolving populations. DNA nucleotides (such as adenine) are not explicitly modeled. With the exception of the evolutionary scenario where we also include beneficial mutations (fig. 8*g*), all new mutations represent transposable elements. Mutations can be of one of two types: "active" or "disabled" transposable elements. All simulations begin with only active transposable elements. Active transposable elements are capable of replication within an individual; active transposable elements replicate with a given probability every generation. When an active transposable element replicates, a new, active transposable element is generated at a random position within that individual's genome. Disabled transposable elements are unable to replicate and are derived from active transposable elements. In all individuals, active transposable elements have a probability of transforming into disabled transposable elements every generation. This would be analogous to a mutation that causes an amino acid change rendering a reverse transcriptase nonfunctional. More information regarding the details of these simulations can be found in our SLiM scripts (https://github.com/gcwoodruff/transposable_elements_2019) and the SLiM manual.

We simulated the transposable element evolutionary dynamics with varying recombination landscapes and fitness effects depending on the genomic region of a transposon element insertion. Chromosomal domains of varying recombination were simulated because such domains are present in *C. elegans* and *C. briggsae* (Rockman and Kruglyak 2009; Ross et al. 2011). Chromosomal domains of varying fitness effects of transposable element insertion were simulated because genes are enriched in chromosome centers in *Caenorhabditis* (supplementary figs. 14 and 15, Supplementary Material online; Cutter et al. 2009). For all simulations, the population size was 5,000 individuals. Genomes were modeled as a single 3 MB chromosome. Recombination was either uniform across the chromosome ($r = 5 \times 10^{-7}$) or had three domains of different recombination rates in the chromosome arms and centers. This recombination rate results in a mode of one crossover event per chromosome per generation in our simulations, consistent with patterns of recombination in *C. elegans* (Rockman and Kruglyak 2009). In the case of three recombination domains, the chromosome arms had high a recombination rate ($r = 5 \times 10^{-7}$), whereas chromosome centers had a lower recombination rate ($r = 1 \times 10^{-9}$). The probability of transposable element replication/insertion was 1×10^{-4}, and the probability of transposable element deactivation was 5×10^{-5}. All simulations were run for 50,000 generations. All scenarios were replicated 50 times. From the simulations, we extracted the number of transposable elements in the central and peripheral domains under different evolutionary scenarios. Additionally, because the birth generation of every transposition event is recorded, we also examined genomic landscapes of transposable element ages in these simulations. All SLiM scripts for simulations have been deposited on Github (https://github.com/gcwoodruff/transposable_elements_2019). We studied the following scenarios:

a. All transposable element insertions are neutral ($s = 0$, fig. 8*a*);
b. Transposable elements are weakly deleterious; the fitness effects are drawn from a gamma distribution with mean $s = -0.0006$ and shape parameter $\alpha = 0.3$ (fig. 8*b*);
c. Transposable elements are highly deleterious; the fitness effects are drawn from a gamma distribution with mean $s = -0.03$ and shape parameter $\alpha = 0.3$ (fig. 8*c*);
d. Transposable elements only located in the central domain are weakly deleterious; fitness effects are drawn from a gamma distribution with mean $s = -0.0006$ and shape parameter $\alpha = 0.3$ (fig. 8*d*);
e. Transposable elements located in the center are more deleterious than in the arms; fitness effects are drawn from a gamma distribution with mean $s = -0.03$ (centers) and mean $s = -0.0006$ (arms) with shape parameter $\alpha = 0.3$ (fig. 8*e*);
f. Transposable elements located in the center are more deleterious than in the arms; fitness effects are drawn from a gamma distribution with mean $s = -0.03$ (centers) and mean $s = -0.0006$ (arms) with shape parameter $\alpha = 0.3$ but transposable elements replicate via a cut-and-paste mechanism. That is, when an element replicates, 50% of the time it is cut and pasted (or moves) into a new position and 50% of the time, it is

copied to a new position (fig. 8*f*). We allow some degree of copying in this case because these transposable elements are always eliminated if they cannot replicate.

g. All transposable elements are weakly deleterious; fitness effects are drawn from a gamma distribution mean $s = -0.0006$ and shape parameter $\alpha = 0.3$. In addition, highly beneficial mutations with fitness effects drawn from a gamma distribution with mean $s = 0.1$ and shape parameter $\alpha = 0.3$ occur with the mutation rate $\mu = 1 \times 10^{-9}$ (fig. 8*g*).

h. All transposable elements have a fixed deleterious fitness effect of insertion. Populations with $s = -0.0002$, $s = -0.0005$, $s = -0.001$, $s = 0.0015$, and $s = -0.002$ were simulated. All simulations with fixed fitness effects of insertion had three domains of different recombination rates along the chromosome (supplementary fig. 28, Supplementary Material online).

i. Transposable elements located in the center are more deleterious than in the arms; fitness effects are drawn from a gamma distribution with mean $s = -0.03$ (centers) and mean $s = -0.0006$ (arms) with shape parameter $\alpha = 0.3$. All individuals reproduce via self-fertilization (supplementary fig. 30, Supplementary Material online).

j. All transposable elements are weakly deleterious; fitness effects are drawn from a gamma distribution with mean $s = -0.03$ (centers) and mean $s = -0.0006$ (arms) with shape parameter $\alpha = 0.3$. All individuals reproduce via self-fertilization (supplementary fig. 31, Supplementary Material online).

Statistical Analyses

All statistical analyses and plots were generated with the R statistical language (R Core Team 2019). The *lm* and *wilcox.test* functions in base R were used for linear models and performing Wilcoxon rank sum tests. The *cohen.d* function in the "effsize" (Torchiano 2020) package was used to estimate Cohen's *d* effect sizes. The "ggplot2" (Wickham 2016), "lemon" (Edwards 2017) and "ggforce" (Pedersen 2019) R packages were used to make plots. Code and data for all statistical analyses have been deposited (https://github.com/gcwoodruff/transposable_elements_2019).

Data Availability

Data files and code associated with this study have been deposited in Github at https://github.com/gcwoodruff/transposable_elements_2019.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Patrick Phillips for sharing genomic data. We thank WormBase and the *Caenorhabditis* genomes project for maintaining and sharing genomic data. Patrick Phillips, Bill Cresko, Peter Ralph, Andy Kern, and their laboratory members provided helpful comments throughout the development of this work. Jeff Adrion provided valuable feedback on earlier versions of this manuscript. This work was supported by funding from the National Institutes of Health Grant to G.C.W. (Grant No. 5F32GM115209-03) and to Patrick Phillips (Grant Nos. R01GM102511, R01AG049396, and R35GM131838).

Author Contributions

G.C.W. remasked genome assemblies, analyzed data, made figures, and wrote the first draft of the paper. A.A.T. performed evolutionary simulations and informed the masking approach. G.C.W. and A.A.T. revised the paper.

References

Adrion JR, Song MJ, Schrider DR, Hahn MW, Schaack S. 2017. Genome-wide estimates of transposable element insertion and deletion rates in *Drosophila melanogaster*. *Genome Biol Evol*. 9(5):1329–1340.

Albertson DG, Thomson JN. 1982. The kinetochores of *Caenorhabditis elegans*. *Chromosoma* 86(3):409–428.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 6:11.

Bao Z, Eddy SR. 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res*. 12(8):1269–1276.

Barnes T, Kohara Y, Coulson A, Hekimi S. 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* 141(1):159–179.

Barrón MG, Fiston-Lavier A-S, Petrov DA, González J. 2014. Population genomics of transposable elements in *Drosophila*. *Annu Rev Genet*. 48(1):561–581.

Bartolomé C, Maside X, Charlesworth B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol Biol Evol*. 19(6):926–937.

Bast J, Schaefer I, Schwander T, Maraun M, Scheu S, Kraaijeveld K. 2016. No accumulation of transposable elements in asexual arthropods. *Mol Biol Evol*. 33(3):697–706.

Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon J-M, Westerman RP, SanMiguel PJ, Bennetzen JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genet*. 5(11):e1000732.

Bégin M, Schoen DJ. 2006. Low impact of germline transposition on the rate of mildly deleterious mutation in *Caenorhabditis elegans*. *Genetics* 174(4):2129–2136.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27(2):573–580.

Bessereau JL. 2006. Transposons in *C. elegans*. *WormBook*. Jan 18:1–13.

Bestor TH. 2000. Sex brings transposons and genomes into conflict. In: McDonald, JF, editor. *Transposable elements and genome evolution*. Dordrecht, Netherlands: Springer. p. 289–295.

Blumenstiel JP, Hartl DL, Lozovsky ER. 2002. Patterns of insertion and deletion in contrasting chromatin domains. *Mol Biol Evol*. 19(12):2211–2225.

Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell*. 24(3):1242–1255.

C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282(5396):2012–2018.

Cabianca DS, Muñoz-Jiménez C, Kalck V, Gaidatzis D, Padeken J, Seeber A, Askjaer P, Gasser SM. 2019. Active chromatin marks drive spatial sequestration of heterochromatin in *C. elegans* nuclei. *Nature* 569(7758):734–739.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*. 10(1):421.

Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Mol Ecol*. 22(6):1503–1517.

Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet*. 23(1):251–287.

Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, et al. 2007. Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317(5836):338–342.

Coghlan A, Tsai IJ, Berriman M. 2018. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protoc Exch*. [10.1038/protex.2018.054] [Internet].

Comeron JM, Ratnappan R, Bailin S. 2012. The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genet*. 8(10):e1002905.

Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol Biol Evol*. 30(10):2311–2327.

Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol*. 25(4):778–786.

Cutter AD. 2015. *Caenorhabditis* evolution in the wild. *BioEssays* 37(9):983–995.

Cutter AD, Dey A, Murray RL. 2009. Evolution of the *Caenorhabditis elegans* genome. *Mol Biol Evol*. 26(6):1199–1234.

Cutter AD, Morran LT, Phillips PC. 2019. Males, outcrossing, and sexual selection in *Caenorhabditis* nematodes. *Genetics* 213(1):27–57.

de Koning AJ, Gu W, Castoe TA, Batzer MA, Pollock DD. 2011. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 7(12):e1002384.

Dey A, Chan CK, Thomas CG, Cutter AD. 2013. Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. *Proc Natl Acad Sci U S A*. 110(27):11056–11060.

Dolgin ES, Charlesworth B. 2008. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* 178(4):2169–2177.

Dubin MJ, Zhang P, Meng D, Remigereau M-S, Osborne EJ, Paolo Casale F, Drewe P, Kahles A, Jean G, Vilhjálmsson B, et al. 2015. DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *Elife* 4:e05255.

Duret L, Marais G, Biémont C. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* 156:1661–1669.

Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460–2461.

Edwards SM. 2017. lemon: freshing up your 'ggplot2' plots. R package version 0.3.1. Available from: https://github.com/stefanedwards/lemon.

Eide D, Anderson P. 1985. Transposition of Tc1 in the nematode *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A*. 82(6):1756–1760.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9(1):18.

Emms DM, Kelly S. 2018. OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. *BioRxiv* [10.1101/466201].

Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, Ahearne TE, Cutter AD, Phillips PC. 2015. Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. *PLoS Genet*. 11(6):e1005323.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 44(D1):D279–D285.

Fischer S, Ruvkun G. 2019. *Caenorhabditis elegans* ADAR editing and the ERI-6/7/MOV10 RNAi pathway silence endogenous viral elements and LTR retrotransposons. *BioRxiv* [825315].

Garrigues JM, Sidoli S, Garcia BA, Strome S. 2015. Defining heterochromatin in *C. elegans* through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. *Genome Res* . 25(1):76–88.

Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* 330(6012):1775–1787.

Glémin S, François CM, Galtier N. 2019. Genome evolution in outcrossing vs. selfing vs. asexual species. In: Anisimova M, editor. *Evolutionary genomics*. New York: Springer. p. 331–369.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453(7197):948–951.

Haller BC, Messer PW. 2019. SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model. *Mol Biol Evol*. 36(3):632–637.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res*. 8(3):269–294.

Hiraki H, Kagoshima H, Kraus C, Schiffer PH, Ueta Y, Kroiher M, Schierenberg E, Kohara Y. 2017. Genome analysis of Diploscapter coronatus: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genomics* 18(1):478.

Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. 2017. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol Biochem Parasitol*. 215:2–10.

Howe M, McDonald KL, Albertson DG, Meyer BJ. 2001. HIM-10 is required for kinetochore structure and function on *Caenorhabditis elegans* holocentric chromosomes. *J Cell Biol*. 153(6):1227–1238.

Ishikawa A, Kabeya N, Ikeya K, Kakioka R, Cech JN, Osada N, Leal MC, Inoue J, Kume M, Toyoda A, et al. 2019. A key metabolic gene for recurrent freshwater colonization and radiation in fishes. *Science* 364(6443):886–889.,

Jenner RA, Wills MA. 2007. The choice of model organisms in evo–devo. *Nat Rev Genet*. 8(4):311–314.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen C-F, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. *Genome Res*. 14(4):528–538.

Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al. 2003. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421(6920):231–237.

Kanzaki N, Tsai IJ, Tanaka R, Hunt VL, Tsuyama K, Liu D, Maeda Y, Namai S, Kumagai R, Tracey A, et al. 2018. Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. *Nat Commun*. 9(1):3216.

Kapusta A, Suh A, Feschotte C. 2017. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A*. 114(8):E1460–E1469.

Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philos Trans R Soc B*. 372(1736):20160458.

Kiontke K, Hironaka M, Sudhaus W. 2002. Description of *Caenorhabditis japonica* n. sp. (Nematoda: Rhabditida) associated with the burrower bug *Parastrachia japonensis* (Heteroptera: Cydnidae) in Japan. *Nematology* 4(8):933–941.

Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, Fitch DH. 2011. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol Biol*. 11:339.

Kofler R, Nolte V, Schlötterer C. 2015. Tempo and mode of transposable element activity in *Drosophila*. *PLoS Genet*. 11(7):e1005406.

Langley CH, Montgomery E, Hudson R, Kaplan N, Charlesworth B. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genet Res*. 52(3):223–235.

Laricchia K, Zdraljevic S, Cook D, Andersen E. 2017. Natural variation in the distribution and abundance of transposable elements across the *Caenorhabditis elegans* species. *Mol Biol Evol*. 34(9):2187–2202.

Lee Y. 2015. The role of piRNA-mediated epigenetic silencing in the population dynamics of transposable elements in *Drosophila melanogaster*. *PLoS Genet*. 11(6):e1005269.

Lee YCG, Langley CH. 2010. Transposable elements in natural populations of *Drosophila melanogaster*. *Philos Trans R Soc B*. 365(1544):1219–1228.

Liu S, Yeh C-T, Ji T, Ying K, Wu H, Tang HM, Fu Y, Nettleton D, Schnable PS. 2009. Mu transposon insertion sites and meiotic recombination events co-localize with epigenetic marks for open chromatin across the maize genome. *PLoS Genet*. 5(11):e1000733.

Lockton S, Gaut BS. 2010. The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol Biol*. 10(1):10.

Lockton S, Ross-Ibarra J, Gaut BS. 2008. Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*. *Proc Natl Acad Sci U S A*. 105(37):13965–13970.

Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.

Medstrand P, Van De Lagemaat LN, Mager DL. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res*. 12(10):1483–1495.

Morgan MT. 2001. Transposable element number in mixed mating populations. *Genet Res*. 77(3):261–275.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310(5746):321–324.

Nowell RW, Almeida P, Wilson CG, Smith TP, Fontaneto D, Crisp A, Micklem G, Tunnacliffe A, Boschetti C, Barraclough TG. 2018. Comparative genomics of bdelloid rotifers: insights from desiccating and nondesiccating species. *PLoS Biol*. 16(4):e2004830.

Okumura E, Tanaka R, Yoshiga T. 2013. Species-specific recognition of the carrier insect by dauer larvae of the nematode *Caenorhabditis japonica*. *J Exp Biol*. 216(4):568–572.

Oliver KR, Greene WK. 2009. Transposable elements: powerful facilitators of evolution. *BioEssays* 31(7):703–714.

Pan J, Sasaki M, Kniewel R, Murakami H, Blitzblau HG, Tischfield SE, Zhu X, Neale MJ, Jasin M, Socci ND, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144(5):719–731.

Pasyukova E, Nuzhdin S, Morozova T, Mackay T. 2004. Accumulation of transposable elements in the genome of *Drosophila melanogaster* is associated with a decrease in fitness. *J Heredity*. 95(4):284–290.

Peacock WJ, Lohe AR, Gerlach WL, Dunsmuir P, Dennis ES, Appels R. 1978. Fine structure and evolution of DNA in heterochromatin. *Cold Spring Harbor Symp Quant Biol*. 42:1121–1135.

Pedersen TL. 2019. ggforce: accelerating 'ggplot2'. R package version 0.3.1. Available from: https://CRAN.R-project.org/package=ggforce

Peng JC, Karpen GH. 2008. Epigenetic regulation of heterochromatic DNA stability. *Curr Opin Genet Dev*. 18(2):204–211.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evol Biol*. 19(1):11.

Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol*. 28(5):1633–1644.

Piatek MJ, Werner A. 2014. Endogenous siRNAs: regulators of internal affairs. *Biochem Soc Trans*. 42(4):1174–1179.

Plohl M, Meštrović N, Mravinac B. 2014. Centromere identity from the DNA point of view. *Chromosoma* 123(4):313–325.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21(Suppl 1):i351–i358.

R Core Team. 2019. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/

Raff RA. 2012. The shape of life: genes, development, and the evolution of animal form. University of Chicago Press.

Ren X, Li R, Wei X, Bi Y, Ho VWS, Ding Q, Xu Z, Zhang Z, Hsieh C-L, Young A, et al. 2018. Genomic basis of recombination suppression in the hybrid between *Caenorhabditis briggsae* and *C. nigoni*. *Nucleic Acids Res*. 46(3):1295–1307.

Rizzon C, Marais G, Gouy M, Biémont C. 2002. Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Res*. 12(3):400–407.

Rizzon C, Martin E, Marais G, Duret L, Ségalat L, Biémont C. 2003. Patterns of selection against transposons inferred from the distribution of Tc1, Tc3 and Tc5 insertions in the *mut-7* line of the nematode *Caenorhabditis elegans*. *Genetics* 165(3):1127–1135.

Rockman MV, Kruglyak L. 2009. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet*. 5(3):e1000419.

Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, Baird SE, Haag ES. 2011. *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet*. 7(7):e1002174.

Schemberger MO, Nascimento VD, Coan R, Ramos É, Nogaroto V, Ziemniczak K, Valente GT, Moreira-Filho O, Martins C, Vicari MR. 2019. DNA transposon invasion and microsatellite accumulation guide W chromosome differentiation in a Neotropical fish genome. *Chromosoma* 128(4):547–560.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, et al. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956): 1112–1115.

Schulenburg H, Félix M-A. 2017. The natural biotic environment of *Caenorhabditis elegans*. *Genetics* 206(1):55–86.

Shapiro JA, Von Sternberg R. 2005. Why repetitive DNA is essential to genome function. *Biol Rev*. 80(2):227–250.

Shi Z, Montgomery TA, Qi Y, Ruvkun G. 2013. High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome Res*. 23(3):497–508.

Shifman S, Bell JT, Copley RR, Taylor MS, Williams RW, Mott R, Flint J. 2006. A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol*. 4(12):e395.

Sijen T, Plasterk RH. 2003. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* 426(6964):310–314.

Slos D, Sudhaus W, Stevens L, Bert W, Blaxter M. 2017. *Caenorhabditis monodelphis* sp. n.: defining the stem morphology and genomics of the genus *Caenorhabditis*. *BMC Zool*. 2(1):4.

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. 2003. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol*. 1(2):e45.,

Steinbiss S, Willhoeft U, Gremme G, Kurtz S. 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. *Nucleic Acids Res*. 37(21):7002–7013.

Stevens L, Félix M-A, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch DH, Frézal L, Kaur T, Kiontke KC, et al. 2019. Comparative genomics of ten new *Caenorhabditis* species. *Evol Lett*. 3(2):217–236.

Stevens L, Rooke S, Falzon LC, Machuka EM, Momanyi K, Murungi MK, Njoroge SM, Odinga CO, Ogendo A, Ogola J. 2020. The genome of *Caenorhabditis bovis*. *Curr Biol*. 30(6):1023–1031.

Stitzer MC, Anderson SN, Springer NM, Ross-Ibarra J. 2019. The genomic ecosystem of transposable elements in maize. *BioRxiv*. [559922].

Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R. 2016. Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation. *Elife* 5:e20777.

Sudhaus W, Giblin-Davis R, Kiontke K. 2011. Description of *Caenorhabditis angaria* n. sp. (Nematoda: Rhabditidae), an associate of sugarcane and palm weevils (Coleoptera: Curculionidae). *Nematology*. 13(1):61–78.

Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. 2016. Genetic drift, not life history or RNAi, determine long-term evolution of transposable elements. *Genome Biol Evol*. 8(9): 2964–2978.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 25(1):4.10.11–14.10.14.

Teterina AA, Willis JH, Phillips PC. 2020. Chromosome-level assembly of the *Caenorhabditis remanei* genome reveals conserved patterns of nematode genome organization.Genetics 214:769 –780.

Torchiano M. 2020. effsize: efficient effect size computation. R package version 0.8.0. Available from: https://doi.org/10.5281/zenodo.1480624

van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534(7605):102–105.

Verma RS. 1988. Heterochromatin: molecular and structural aspects. New York: Cambridge University Press.

Volk J. 1951. Die Nematoden der Regenwurmer und aasbesuchenden Kafer. *Zool Jb Syst*. 79:1–70.

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, et al. 2007. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 8(12):973–982.

Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer.

Woodruff GC, Johnson E, Phillips PC. 2019. A large close relative of *C. elegans* is slow-developing but not long-lived. *BMC Evol Biol*. 19(1):74.

Woodruff GC, Phillips PC. 2018. Field studies reveal a close relative of *C. elegans* thrives in the fresh figs of Ficus septica and disperses on its Ceratosolen pollinating wasps. *BMC Ecol*. 18(1):26.

Woodruff GC, Willis JH, Phillips PC. 2018. Dramatic evolution of body length due to post-embryonic changes in cell size in a newly discovered close relative of *C. elegans*. *Evol Lett*.

Wright SI, Agrawal N, Bureau TE. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res*. 13(8):1897–1903.

Wright SI, Le QH, Schoen DJ, Bureau TE. 2001. Population dynamics of an Ac-like transposable element in self-and cross-pollinating *Arabidopsis*. *Genetics*. 158(3):1279–1288.

Wright SI, Schoen DJ. 1999. Transposon dynamics and the breeding system. *Genetica* 107(1/3):139–148.

Yigit E, Batista PJ, Bei Y, Pang KM, Chen C-C, Tolia NH, Joshua-Tor L, Mitani S, Simard MJ, Mello CC. 2006. Analysis of the *C. elegans* Argonaute family reveals that distinct Argonautes act sequentially during RNAi. *Cell* 127(4):747–757.

Yin D, Schwarz EM, Thomas CG, Felde RL, Korf IF, Cutter AD, Schartner CM, Ralston EJ, Meyer BJ, Haag ES. 2018. Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science* 359(6371):55–61.

Yoshiga T, Ishikawa Y, Tanaka R, Hironaka M, Okumura E. 2013. Species-specific and female host-biased ectophoresy in the roundworm *Caenorhabditis japonica*. *Naturwissenschaften* 100(2):205–208.