Pervasive Conservation of Intron Number and Other Genetic Elements Revealed by a Chromosome-level Genome Assembly of the Hyper-polymorphic Nematode *Caenorhabditis brenneri*

Anastasia A. Teterina [1,2], John H. Willis [1], Charles F. Baer [3], Patrick C. Phillips [1,*]

[1]Institute of Ecology and Evolution, University of Oregon, Eugene, OR, USA
[2]Center of Parasitology, Severtsov Institute of Ecology and Evolution RAS, Moscow, Russia
[3]Department of Biology, University of Florida, Gainesville, FL, USA

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

Accepted: February 25, 2025

Abstract

With within-species genetic diversity estimates that span the gamut of that seen across the entirety of animals, the *Caenorhabditis* genus of nematodes holds unique potential to provide insights into how population size and reproductive strategies influence gene and genome organization and evolution. Our study focuses on *Caenorhabditis brenneri*, currently known as one of the most genetically diverse nematodes within its genus and, notably, across Metazoa. Here, we present a high-quality, gapless genome assembly and annotation for *C. brenneri*, revealing a common nematode chromosome arrangement characterized by gene-dense central regions and repeat-rich arms. A comparison of *C. brenneri* with other nematodes from the "*Elegans*" group revealed conserved macrosynteny but a lack of microsynteny, characterized by frequent rearrangements and low correlation of orthogroup size, indicative of high rates of gene turnover, consistent with previous studies. We also assessed genome organization within corresponding syntenic blocks in selfing and outcrossing species, affirming that selfing species predominantly experience loss of both genes and intergenic DNA. A comparison of gene structures revealed a strikingly small number of shared introns across species, yet consistent distributions of intron number and length, regardless of population size or reproductive mode, suggesting that their evolutionary dynamics are primarily reflective of functional constraints. Our study provides valuable insights into genome evolution and expands the nematode genome resources with the highly genetically diverse *C. brenneri*, facilitating research into various aspects of nematode biology and evolutionary processes.

Key words: *Caenorhabditis* nematodes, genome assembly and annotation, comparative genomics, genome organization, exon/intron gene structure.

Significance

Effective population size is thought to be a primary driver of genomic evolution across all organisms. *Caenorhabditis* nematodes offer a unique opportunity to test these ideas because their effective population sizes span many orders of magnitude. We generated a complete genome of *Caenorhabditis brenneri*, one of the most diverse metazoans known, and compared patterns of genomic organization across species with varying levels of diversity. A comparison of gene structures shows few shared introns but conserved distributions of intron number and length, regardless of population size or reproductive mode, which suggests that these elements are under strong functional constraint. Overall, these patterns are consistent with strong selection on genomic features often considered to be subject to neutral genetic drift.

Introduction

A central tenet of molecular evolution is that the pace and structure of the evolution of genomes and genetic elements should be largely determined by variation in population size. This is because it is thought that the expansion of some genetic elements, such as repetitive DNA and introns, should be weakly deleterious (Carmel and Chorev 2012; Burns 2022), and therefore be expected to grow in small populations ($4N_es < 1$, where N_e is the effective population size and s is the selection coefficient) and be eliminated in large populations ($4N_es > 1$) through a balance of natural selection and genetic drift. For example, Lynch and colleagues have suggested that a major determinant of differences in genome and gene structures among prokaryotes and eukaryotes is due to differences in effective population size (Lynch 2002; Lynch 2007; however, see discrepancies in Charlesworth and Barton 2004; Daubin and Moran 2004; Vinogradov 2004). Within animals, there tends to be some signature of the influence of population size on genome size and structure (Gregory et al. 2007; Galtier 2024), but the results are often equivocal (Roddy et al. 2021; Marino et al. 2025). Often vast evolutionary distances are included in these comparisons, which can lead to challenges of phylogenetic nonindependence in the analyses (Whitney and Garland 2010; Whitney et al. 2011), as well as the likelihood that the actual functional role of genetic elements of interest may themselves have changed. Nematodes of the family *Caenorhabditis* hold the promise to circumvent some of these challenges, as species within this group include some of the least diverse and most diverse animals known (Dey et al. 2013; Noble et al. 2021), indicating vast differences in effective population size. Until recently, however, complete genomes have only been available for a few selfing species, including the well-studied model system *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium 1998; Stein et al. 2003; Noble et al. 2021). Genome assemblies of outcrossing species have frequently been compromised by residual polymorphism, although this has been addressed using long-read sequencing in several species with moderate to high levels of within-species polymorphism (Barrière et al. 2009; Kanzaki et al. 2018; Teterina et al. 2020). Here, we push the limits of this comparative framework by presenting a whole-chromosome assembly of one of the most polymorphic animals currently known (Dey et al. 2013; Alm Rosenblad et al. 2021), *Caenorhabditis brenneri*, and use this new assembly in a comparative analysis of the evolution of size and structure of a variety of genetic elements thought to be subject to nearly neutral evolution, particularly introns.

Caenorhabditis nematodes have compact genomes, typically 80 to 140 Mb, with six holocentric chromosomes—five autosomes and one sex chromosome (Brenner 1974; Pires-daSilva 2007; Sun et al. 2022). While the majority of *Caenorhabditis* species have males and females (dioecy) and reproduce via outcrossing, three species have independently shifted to self-reproduction, or "selfing," with most individuals being hermaphrodites with a few males (androdioecy, Kiontke et al. 2011). This mode of reproduction has greatly reduced the effective population size of these species (Golding and Strobeck 1980; Pollak 1987; Nordborg and Donnelly 1997; Thomas et al. 2012b; Cutter 2019). Genomes of outcrossing species within the *Elegans* group are larger and contain a greater number of genes than selfing species (Bird et al. 2005; Fierst et al. 2015; Yin et al. 2018; Stevens et al. 2019; Adams et al. 2023), with some of the reduction being attributable to the shift to self-reproduction per se via resulting changes in the regulation of male-specific pathways (Thomas et al. 2012a, 2012b; Yin et al. 2018; Sánchez-Ramírez et al. 2021; Xie et al. 2022). Nematodes have a conservative chromosome organization, with high gene density in the central parts of chromosomes and repetitive elements concentrated in peripheral "arms" (The *C. elegans* Sequencing Consortium 1998; Carlton et al. 2022). Additionally, they exhibit conservative macro-synteny across *Rhabditid* species, with most orthologous genes consistently located on homologous chromosomes (Tandonnet et al. 2019; De La Rosa et al. 2021). Despite this, *Caenorhabditis* nematodes display a relatively high rate of gene turnover, involving both the expansion and reduction of gene families (Fierst et al. 2015; Adams et al. 2023; Ma et al. 2024). Moreover, the exon/intron structure in orthologous genes in nematodes is not conservative, particularly via the substantial loss of introns (Kent and Zahler 2000; Cho et al. 2004; Coghlan and Wolfe 2004; Kiontke et al. 2004; Stevens et al. 2019; Ma et al. 2022).

Caenorhabditis nematodes are especially useful for testing hypotheses regarding the impact of population size on gene and genome organization because of the very substantial differences in population sizes across species. Namely, outcrossing nematode *C. brenneri*, named in honor of Sydney Brenner for his pivotal role in *C. elegans* research (Sudhaus and Kiontke 2007), has an effective population size of about 10^7, making its genetic diversity on the order of many bacterial species, with nearly one in ten nucleotides being polymorphic between any two individuals and an average diversity of 14.1% at synonymous sites (Dey et al. 2013). Another outcrossing nematode *C. remanei* displays an effective population size of around 10^6 (Graustein et al. 2002; Cutter et al. 2006; Teterina et al. 2023), whereas selfing species like *C. elegans*, *C. briggsae*, and *C. tropicalis* exhibit smaller population sizes, $\sim 10^4$ (based on diversity estimations from Sivasundar and Hey 2003; Barrière and Félix 2005; Cutter 2006; Dolgin et al. 2007; Dolgin et al. 2008; Noble et al. 2021 considering the mutation rate from Saxena et al. 2019). These differences translate into

substantial differences in the probability of fixation of deleterious variation in introns (Fig. 1, Lynch 2002 c.f. Eq. 1a). If assumptions underlying these expected outcomes hold, then we would predict that we should see noticeable disparities in intron size distributions across nematode species. To effectuate the test of this prediction, we generated a gapless chromosome-scale genome assembly and high-quality annotation for *C. brenneri* and assessed divergence and conservation of the genomic landscape of genetic elements across the genus using a comparative framework. As might be anticipated, no single simple factor explains the overall patterns of genomic evolution that we observe, which are characterized by macro-level conservation and micro-level evolutionary dynamics.

Results

Chromosome-scale Genome Assembly of *C. brenneri*

Using 90 generations of brother-sister mating, we generated a highly inbred strain, CFB2252 of *C. brenneri* with greatly reduced residual heterozygosity. Using a combination of long- and short-reads, Hi-C data, and high-accuracy long-read (HiFi) data from an individual nematode, we assembled and scaffolded a gapless genome for *C. brenneri*. Interestingly, the genome exhibited substantial continuity following the initial assembly with long CLR PacBio reads (supplementary fig. S1, Supplementary Material online), likely due to the low level of genetic variation in the CFB2252 strain (residual *k*-mer based heterozygosity of 0.8%), with the entire genome being fully assembled with the inclusion of Hi-C data (supplementary fig. S1, Supplementary Material online). Also, we used HiFi reads to generate a mitochondrial genome (supplementary fig. S2, Supplementary Material online), which matches the previously available mitochondrial genome (NC_035244.1). The HiFi data were generated from a single nematode, with high genome coverage and low proportion of alternative alleles throughout the genome, as expected for an inbred strain (supplementary fig. S3, Supplementary Material online). After removal of the assembly artifacts and bacterial contamination, six gapless chromosomal scaffolds remained, together with 163 additional scaffolds. Some of the latter may be the result from misassembles or alternative structural haplotypes, as they mostly map to the repetitive regions. These additional elements were included in the assembly as unplaced scaffolds. Our new *C. brenneri* assembly shows tremendous improvement if quality and reduction in the duplication level in Benchmarking Universal Single-Copy Orthologs (BUSCO) genes from ~30% to 1% as compared to the previously available assembly (Table 1).

The genome organization of *C. brenneri* exhibits a domain-like structure similar to several other *Caenorhabditis* species (The *C. elegans* Sequencing Consortium 1998; Woodruff and Teterina 2020; Carlton et al. 2022) with more repeats and lower gene density in the peripheral arms of the

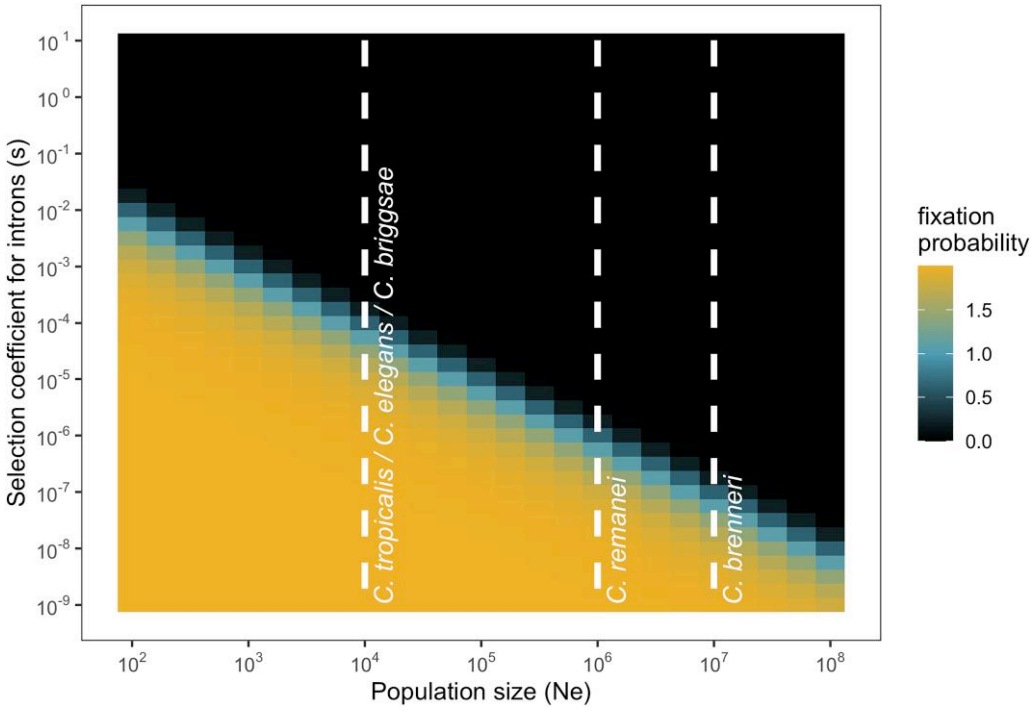

Fig. 1. Scaled probabilities of fixation of introns in diploid populations. The values are based on Equation 1a from Lynch (2002) and are scaled by 2Ne. The white dashed lines demonstrate the order of population sizes for some *Caenorhabditis* nematodes.

Table 1 Assembly statistics for *C. brenneri* assemblies

Caenorhabditis brenneri genome assembly	This study	The *Caenorhabditis brenneri* sequencing and analysis consortium
Accession number	GCA_964036135.1	GCA_000143925.2
Strain	CFB2252	PB2801
Total ungapped length, bp	123,187,963	170,093,652
Percent gaps, %	0	10
Number of scaffolds	6 chromosomes + mtDNA (+163 unplaced scaffolds)	3,305
Scaffold N50	20 Mb	382 kb
Scaffold L50	3	120
GC%	38.42	38.5
QV	36	-
Number of protein-coding genes	25,756	30,670
BUSCO Metazoa, 954 genes	Completness: 71.8% [Single-copy: 70.6%, duplicated: 1.2%] Fragmented: 3.7% Missing: 24.5%	Completness: 72.5% [Single-copy: 45.6%, duplicated: 26.9%] Fragmented: 3.5% Missing: 24.0%
BUSCO Nematoda, 3131 genes	Completness: 97.3% [Single-copy: 96.6%, duplicated: 0.7%] Fragmented: 1.1% Missing: 1.6%	Completness: 97.2% [Single-copy: 65.4%, duplicated: 31.8%] Fragmented: 1.0% Missing: 1.8%

BUSCO statistics were estimated for the databases version 10 using all scaffolds.

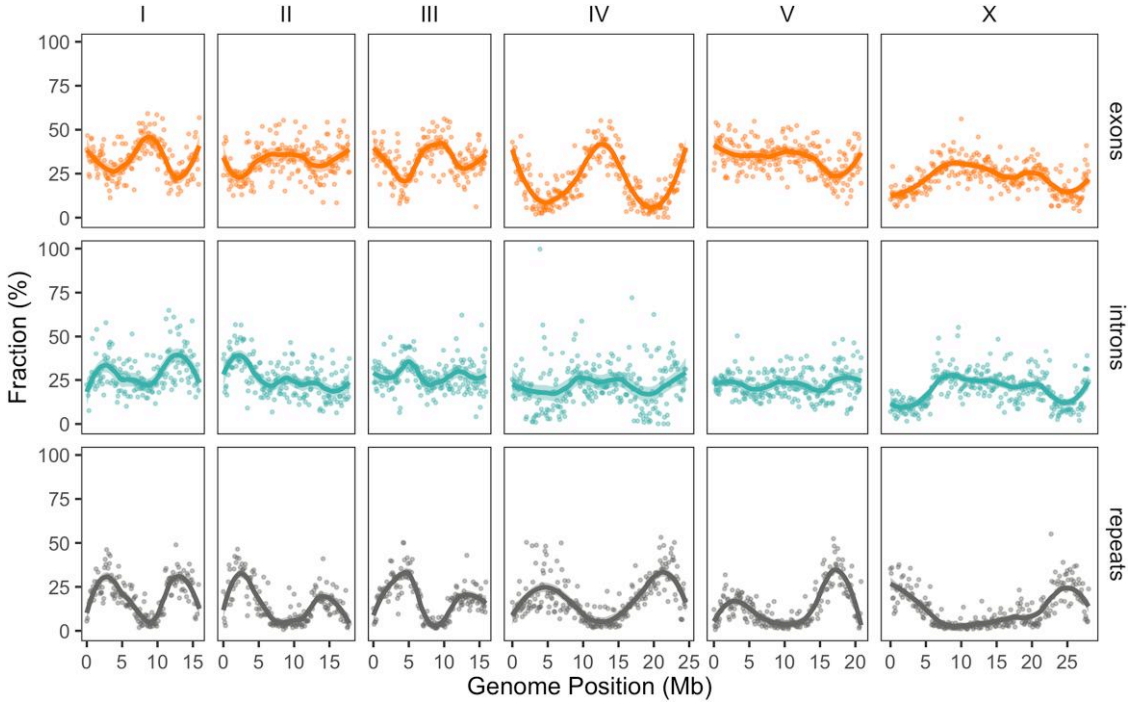

Fig. 2. Genomic landscape of genetic elements for *C. brenneri*. Colored lines represent the smoothed means of the fraction of repetitive DNA, exons, and introns calculated from 100 kb windows. A locally weighted smoothing curves were fitted using with a span of 0.4, shaded areas show 95% CIs of the mean.

chromosomes and greater gene density in central regions across all chromosomes (Fig. 2). Notably, *C. brenneri* shows a relatively low percentage of repeats (only 15.9%), which is lower than other species of *Caenorhabditis* of the *Elegans* group (Fierst et al. 2015; Woodruff and Teterina 2020). This difference may be attributed to the divergence of the repeats from one another, as the repeat annotation protocol we used only masks repeats that are 80% identical.

Central Domains of Chromosomes are More Conserved

To examine the conservation pattern along the genome, we aligned the *C. brenneri* genome with the genomes of several closely related *Caenorhabditis* nematodes. The phylogenetic trees, built from 50 kb windows alignments with more than 1 kb nucleotides aligned across species (mean alignment length: 7,335 ± 5,541 kb), support a sister-species relationship of *C. brenneri* with *C. sp48* throughout the genome (Fig. 3a and b). Only a few windows displayed a different tree topology; however, they were primarily located in the most divergent regions where reconstruction is likely less reliable (Fig. 3a). The tree heights, as well as the lengths of *C. brenneri* and *C. sp48* clade branches, were longer in the peripheral parts of chromosomes (two top rows in Fig. 3c), indicating greater conservation in the central regions of chromosomes. This observation in agreement with previous findings in nematodes (The *C. elegans* Sequencing Consortium 1998; Hillier et al. 2007; Teterina et al. 2020). Additionally, we calculated the mean conservation score for each 50 kb window alignment using the phastCons algorithm (Siepel et al. 2005), a phylogenetic hidden Markov model-based method that estimates the probability of each nucleotide being under negative selection and belonging to conservative elements based on sequence alignments. The central regions exhibited elevated conservation scores (Fig. 3c, bottom row), consistent with the distribution of tree heights. This pattern of central-domain conservation is likely associated with the high gene density as well as the presence of genes with more conserved function being located in the central parts of chromosomes (Parkinson et al. 2004; Cutter et al. 2009). Further, suppressed recombination in the central-domain, which has been observed in several other closely related species (Rockman and Kruglyak 2009; Ross et al. 2011; Teterina et al. 2023), may also contribute to this pattern. When combined with natural selection, these factors can drive down local diversity and divergence through the action of linked selection (Smith and Haigh 1974; Charlesworth et al. 1993; Nordborg et al. 1996; Coop and Ralph 2012).

To examine the pattern of chromosome rearrangements and the evolution of genome organization, we utilized orthogroups and conservative gene order (Lovell et al. 2022) to infer syntenic blocks across *C. brenneri* and six other *Caenorhabditis* species with chromosome-scale assemblies (Fig. 4). *Caenorhabditis* nematodes have six holocentric chromosomes that show almost no interchromosomal rearrangements but exhibit a significant number of intrachromosomal translocations and inversions. *Caenorhabditis nigoni* and *C. briggsae*, closely related species capable of hybridizing with one another (Woodruff et al. 2010), displayed only a few rearrangements. Statistics on the number and sizes of syntenic blocks relative to *C. brenneri* are provided in (supplementary table S1, Supplementary Material online). Naturally, *C. brenneri* had the smallest number and longest syntenic blocks with the most closely related species, *C. tropicalis*. The number of blocks in reversed orientation was noticeably high, ~50%; however, the mean lengths of blocks in reverse orientation tend to be smaller. Repression of interchromosomal rearrangements in nematodes and the stability of karyotypes have been described in other work (Stein et al. 2003; Hillier et al. 2007; Kanzaki et al. 2018; Stevens et al. 2020; Teterina et al. 2020; De La Rosa et al. 2021; Sun et al. 2022) and could potentially be a consequence of large population sizes and selection against large interchromosomal translocations, meiotic control mechanisms such as pairing centers (McKim et al. 1993; Villeneuve 1994; reviewed in Carlton et al. 2022), and/or other mechanisms controlling genome stability during meiosis (Hillers and Villeneuve 2003; Lui and Colaiácovo 2013; Bhargava et al. 2020; Dokshin et al. 2020).

Genome Organization in Selfing and Outcrossing Species

Self-reproduction can affect genomic organization due to higher genetic drift relative to outcrossing species. It has been demonstrated that selfing nematodes in the *Elegans* group have 10% to 30% smaller genomes than outcrossing ones, predominately due to gene loss (Fierst et al. 2015; Yin et al. 2018; Adams et al. 2023), balanced against possible gene family expansion in outcrossing species (Kanzaki et al. 2018; Stevens et al. 2019; Adams et al. 2023). Subsequent to these studies, additional chromosome-scale genomes for outcrossing species of the *Elegans* group have become available (Kanzaki et al. 2018; Teterina et al. 2020; and this work), enabling us to investigate specific features of genome organization that distinguish selfing from outcrossing species. We employed both generalized linear models and gradient boosting, a machine learning method based on ensemble learning with decision trees, to classify species as either outcrossing or selfing. The classification criteria included ratios of fractions and numbers of exons, introns, and genes, as well as the size of the blocks and intergenic regions within a species, relative to the outcrossing species *C. brenneri* (supplementary fig. S4, Supplementary Material online). These comparisons were made within inferred syntenic blocks, which were replicated proportionally to their size in *C. brenneri*. The models were trained on two selfing species, *C. elegans* and *C. tropicalis*, and two outcrossing species, *C. inopinata* and *C. remanei*. Subsequently, we tested the models on ratios from two closely related species, selfing *C. briggsae* and *C. nigoni*. The results from the generalized linear model (GLM) were more accurate than those from the gradient boosting model approach (supplementary fig. S5, Supplementary Material online). We were able to successfully classify selfing and outcrossing species. The fraction of intergenic DNA and exon

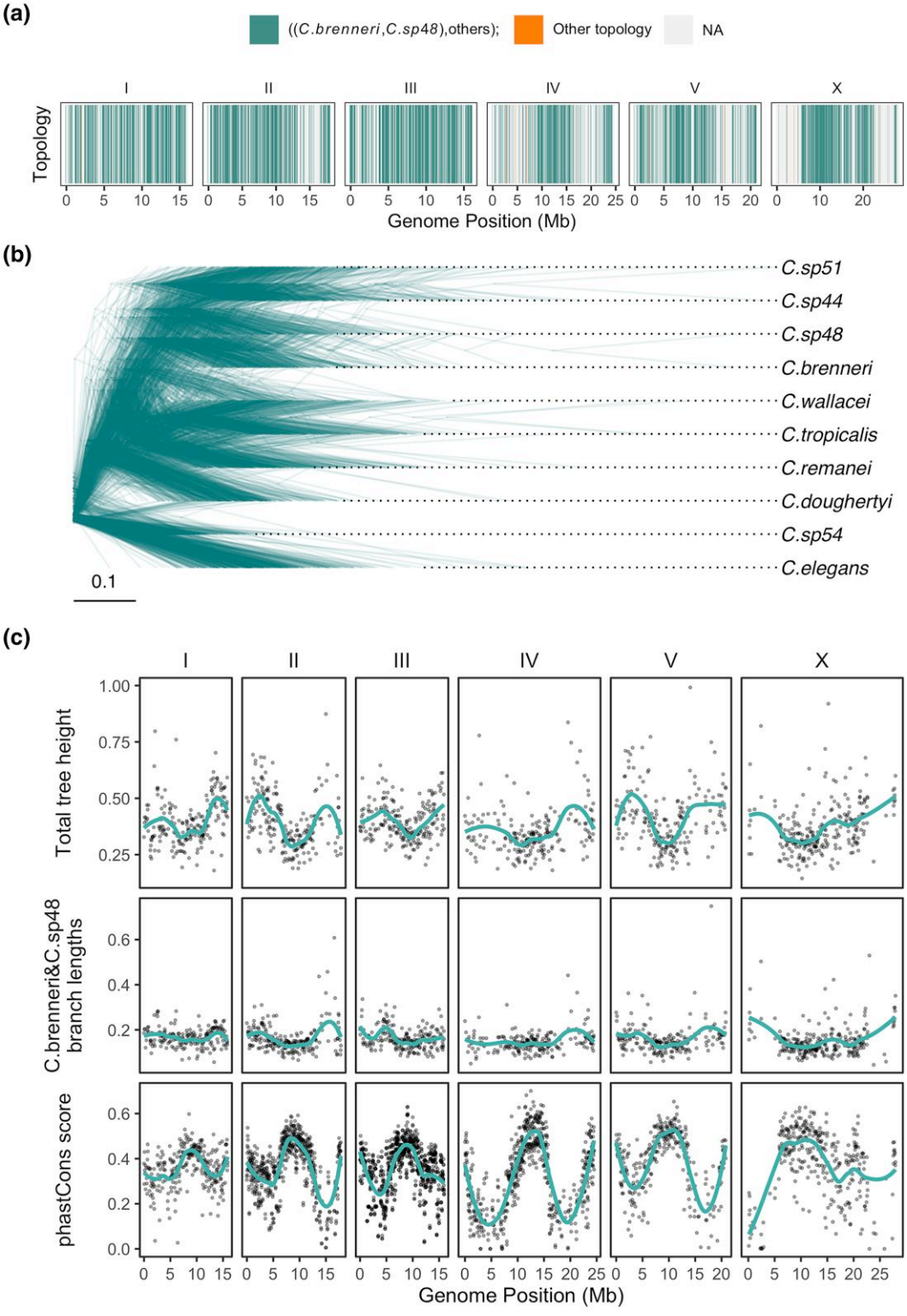

Fig. 3. Divergence landscape and phylogeny of *C. brenneri* in 50 kb nonoverlapping genomic windows, with alignments longer than 1 kb across all species. A locally weighted smoothing curves were fitted using with a span of 0.4. a) The tree topology along the genome of *C. brenneri*. Teal color indicates that *C. brenneri* is the sister species with *C. sp48* (1303 windows), while orange represents other topologies (12 windows), and light gray indicates missing data (1151 windows). b) Maximum-likelihood trees sampled across the genome (500 trees). c) Tree statistics along the genome of *C. brenneri*. The top row displays the total tree heights; the middle row illustrates the length of the *C. brenneri* and *C. sp48* branch. Conservation scores along the genome, estimated using the phastCons algorithm, are shown in the bottom row.

Fig. 4. The riparian plot shows syntenic blocks among *Caenorhabditis* nematodes, based on the order of orthologues genes, demonstrating substantial intrachromosomal rearrangements. The central domains exhibit more extended syntenic blocks. *Caenorhabditis brenneri* was used as a reference. Also, chromosome IV of *C. remanei* was rotated.

count were the most important features for classification (Fig. 5a). Figure 5b shows the coefficients and significance for various factors and their interactions used in the GLM, and Fig. 5c displays the permutation importance of those factors. The results support prior findings, demonstrating that selfing species lose a considerable number of genes but also a substantial amount of intergenic noncoding DNA (Fierst et al. 2015; Yin et al. 2018; Cutter et al. 2019; Adams et al. 2023). More extensive annotation of various functional and regulatory elements should allow these discrepancies to be explored in greater detail.

Influence of Population Size on Gene Organization

To explore the exon/intron structure of genes in nematodes, we compared orthologous genes among seven *Caenorhabditis* species with chromosome-scale assemblies. Nematodes display a low correlation in orthogroup sizes that rapidly decays with phylogenetic distance and is significantly associated with it (Fig. 6a; Mantel test using 1—correlation, 999 permutations: Z-score = 6.025, P-value = 0.003). This observation is in accordance with data indicating high rates of gene turnover (Adams et al. 2023; Ma et al. 2024). However, when estimating the difference in sizes of shared orthogroups, the correlations were close to zero (supplementary fig. S6A, Supplementary Material online), and the percentage of shared orthogroups varies among species pairs. (Supplementary fig. S6b, Supplementary Material online) illustrates the percentage of shared orthogroups and the total number of genes in them. Outcrossing nematodes have on average 2.4 times more species-specific orthogroups than selfing species. We subsequently aligned exon/intron structures of 7,238 single-copy (hereinafter, 1-to-1) orthogroups. Several examples of such alignments are demonstrated in (supplementary fig. S7, Supplementary Material online). There, the majority of coding sequences are conserved and present in all species, but the positions and lengths of introns vary greatly, even as the number of introns remains mostly consistent. Correlations of gene structure between species pairs were low (Fig. 6b), but significantly associated with phylogenetic distance (Mantel test using 1—correlation with 999 permutations: Z-score = 6.834, P-value = 0.005). Only 1% of introns are shared among all species (see the inner plot in Fig. 6b), and 64.3% of the 167,873 introns from single-copy orthologs are unique. While several previous studies have demonstrated a high rate of intron loss in nematodes (Kent and Zahler 2000; Cho et al. 2004; Coghlan and Wolfe 2004; Kiontke et al. 2004), others focusing on highly conservative genes have reported a much higher level of conservation in exon/intron structure (Ma et al. 2022). Our work, which includes a small set of species that have nearly complete genomes, reveals very low levels of gene structure conservation, and that even within an examination of a small fraction of known Clade V nematodes.

We compared gene, exon, and intron length distributions, as well as the number of exons, among both all genes and

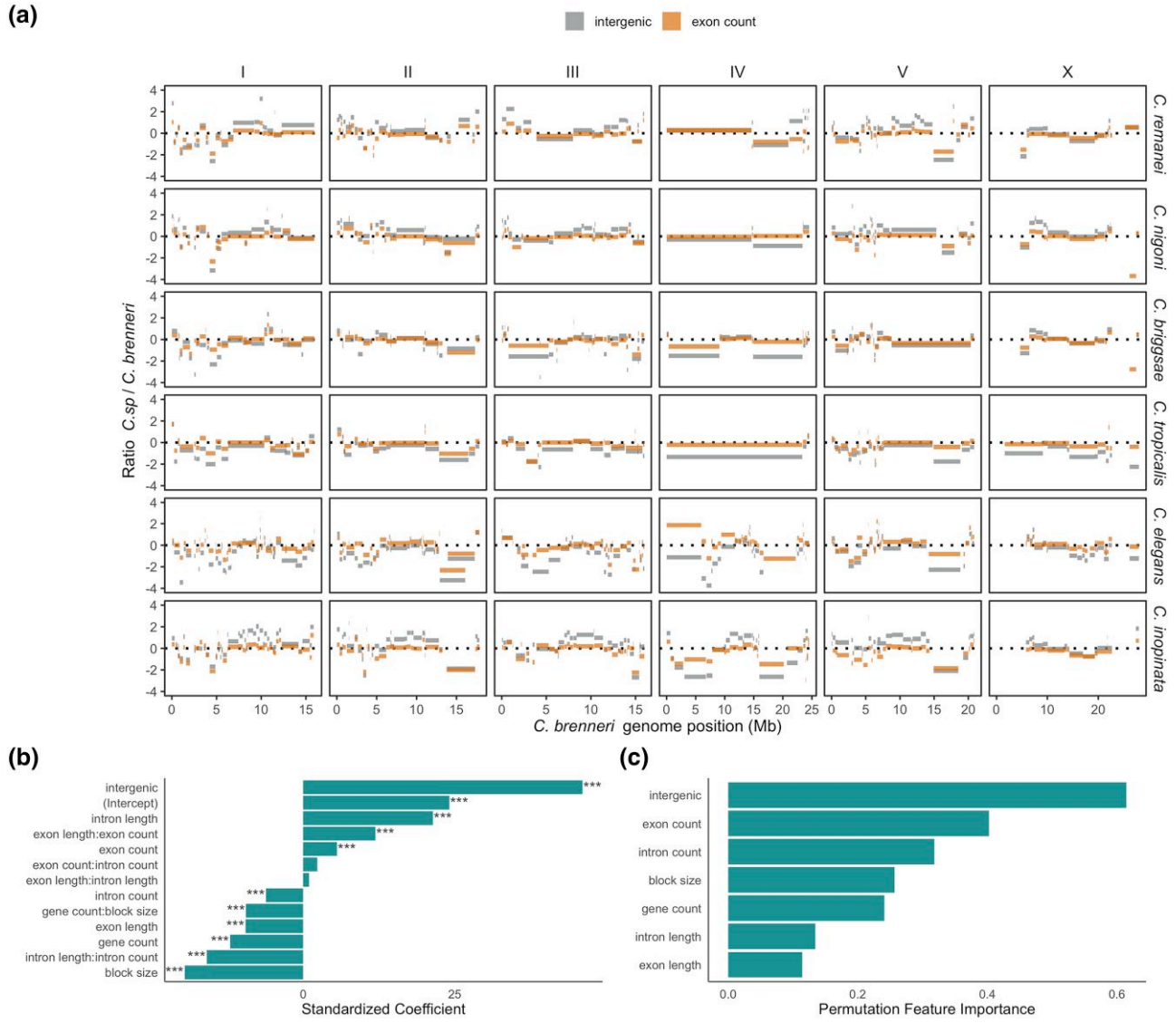

Fig. 5. Comparison of genetic elements in *Caenorhabditis* species relative to *C. brenneri* and classification of selfing and outcrossing species using those features. a) Ratios of exon counts (orange) and fraction of intergenic regions (gray) between a focal species and *C. brenneri* in corresponding syntenic blocks (see Fig. 4) along the *C. brenneri* genome. b) Coefficients of genomic features used in the GLM for classifying selfing and outcrossing species. *** denotes Bonferroni-corrected *P*-values < 0.0001. c) Permutation importance of features for classifying nematodes into selfing and outcrossing species.

single-copy orthologues genes in *Caenorhabditis* nematodes. Figure 7a illustrates the mean and median values of exon and intron lengths, and (supplementary fig. S8, Supplementary Material online) the gene lengths for corresponding species, with the effect sizes (Cohen's *d*) and significance from the nonpaired Wilcoxon tests in 1-to-1 orthologues provided in (supplementary fig. S9, Supplementary Material online). Overall, these seven nematodes have very similar distributions in sizes of genes, exons, and introns and the number of exons (Fig. 7b, see also Stevens et al. 2019), despite substantial differences in effective population size across species. Further, all species have more introns in 1-to-1 orthologs than the average observed across all genes. *Caenorhabditis elegans* and *C. inopinata* tend to have longer genes and introns (supplementary fig. S9a and c, Supplementary Material online), and *C. nigoni* has fewer introns than other species (supplementary fig. S9b, Supplementary Material online). The quality of annotations in these species may impact some statistical results; however, because we used only fully assembled well-annotated genomes and because 1-to-1 orthologs tend to be more conserved, we expect the analysis to be reliable.

Differences between selfing (*C. elegans*, *C. briggsae*, *C. tropicalis*) and outcrossing species (*C. brenneri*, *C. remanei*, *C. nigoni*, *C. inopinata*) in those gene features are summarized in Table 2. When comparing all genes,

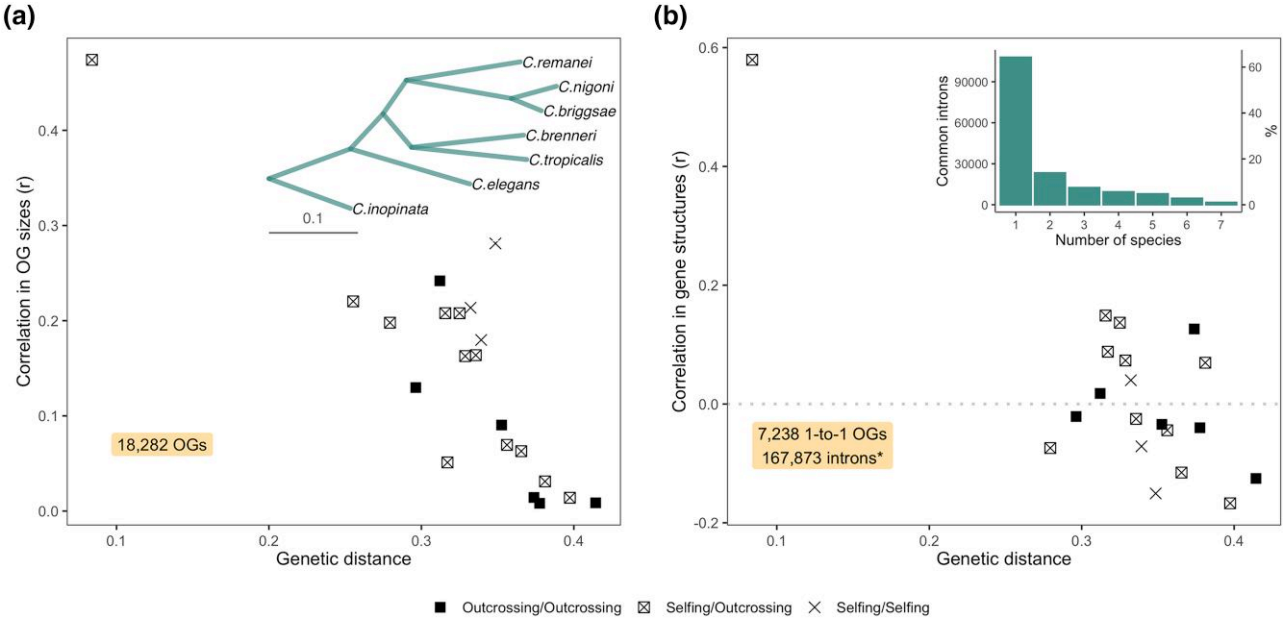

Fig. 6. Correlations of orthogroup sizes and gene structures between species pairs, plotted against genetic distances. a) The correlation of orthogroup sizes across 18,282 orthogroups against the phylogenetic distance between species, using only orthogroups present in both species. The inset plot displays the phylogenetic tree from orthoFinder analysis, used for estimating genetic distances. Black boxes show comparisons of outcrossing species versus outcrossing species, boxes with crosses represent selfing versus outcrossing species, and crosses signify selfing versus selfing species. b) The correlation of exon/intron gene structure across 7,238 single-copy orthologs present in all species. * We used 167,873 introns determined by GenePainter analysis, which includes only introns surrounded by coding sequences, without 5′ UTR introns. The inset plot shows a histogram illustrating the frequency of common introns across nematode species, where "1" represents species-specific introns.

selfing species tend to have longer genes and have approximately one more intron than outcrossing species. However, there are no discernible distinctions in average gene, exon, and intron lengths, as well as exon numbers, when comparing these features in 1-to-1 orthologs across species with selfing and outcrossing reproductive modes and, thus, drastically distinct population sizes.

We used the alignments of exon/intron structures of 1-to-1 orthologs to compare certain features of shared introns, including length, phase, and splice sites, across pairs of species, and between selfing and outcrossing nematodes. Interestingly, sister species *C. elegans* and *C. inopinata* on average have longer introns than other species (supplementary fig. S10a, Supplementary Material online). Using a linear mixed-effects model, we find no evidence that phylogenetic distance or reproductive mode has a substantial influence on the length differences of shared introns.

In addition to intron size, we compared the phases and splice sites of introns shared across nematode species, as well as those unique within each species. Phase refers to the position of the intron relative to the three based of the codon (e.g. phase 0 means that the intron begins precisely at the end of the preceding codon). All species have more than 50% of introns in phase 0 (Fig. 8a), consistent with the general observation that phase 0 U2 introns are the most frequent in most organisms (Fedorov et al. 1992; Long et al. 1995; Nguyen et al. 2006; Rogozin et al. 2012), though the functional mechanisms behind this pattern remain unclear (Fedorov et al. 1992; de Souza et al. 1998; Olthof et al. 2024). *Caenorhabditis elegans*, *C. briggsae*, *C. inopinata*, and *C. nigoni* showed a significant difference (Fisher's test) in the distribution of introns between phase 0 and nonphase 0 (combining phase 1 and phase 2) for the species-specific introns ("Unique" in Fig. 8). In contrast, *C. brenneri* exhibited a significant difference between species-specific and common introns, with more nonphase 0 introns in those shared with other species. The majority of splice sites of introns in 1-to-1 orthologs are the canonical GT-AG sites. We also compared variation in the fraction of other rare splice sites between species-specific and shared introns among nematodes (Fig. 8b), find no significant differences for the majority of species (Fisher's test), except for the selfing *C. briggsae*, which exhibited more frequent rare types of splice sites in species-specific introns (analysis of phase-based data can be found in supplementary fig. S10b and c in Supplementary material online).

Discussion

Nematodes provide a valuable metazoan model to study the impact of effective population size and reproductive

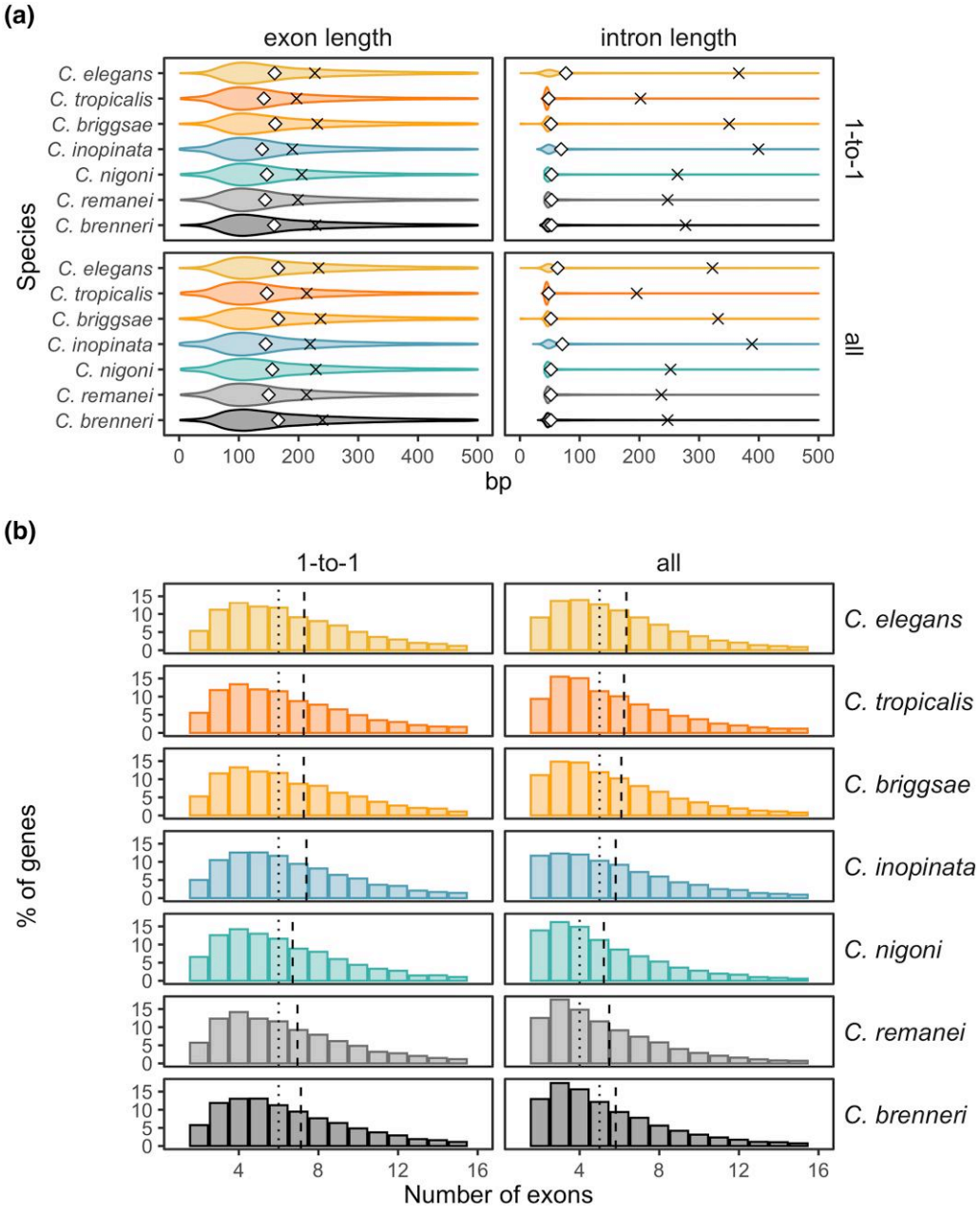

Fig. 7. Distributions of genetic features in all genes and single-copy orthologs. a) Exon and intron lengths in single-copy (1-to-1) orthologs and all genes within the orthogroups of the species. The crosses represent the mean values, and the diamonds represent the medians. b) The number of exons in 1-to-1 orthologs and all genes within orthogroups. The dashed vertical lines indicate the means, and the dotted vertical lines indicate the medians.

system on gene and genome organizations. *Caenorhabditis brenneri* is currently considered the most genetically diverse among all *Caenorhabditis* nematodes, and among the most diverse known metazoans (Dey et al. 2013), and completion of a high-quality genome and annotation for this species greatly expands the total scope of variation across this genus. The chromosome organization of *C. brenneri* follows the typical pattern observed in nematodes, featuring conservative macrosynteny across species, and large central domains with high gene density and low repeat density that exhibit more conservation than peripheral domains. Selfing species have lost more protein-coding and noncoding DNA relative to outcrossing species; however, all seven *Caenorhabditis* species have low correlation in orthogroup sizes and very high rates of gene turnover. Somewhat surprisingly, given the general pattern of synteny seen at a macro scale, gene exon/intron structure at the individual gene level varies substantially between the seven analyzed

Table 2 Differences in gene features in selfing and outcrossing species

	Selfing versus outcrossing species					
	1-to-1 orthologs			All genes		
	Mean ± SD in selfers; outcrossers	Cohen's *D*	Linear model *P*-value, Bonferroni correction	Mean ± SD in selfers; outcrossers	Cohen's *D*	Linear model *P*-value, Bonferroni correction
Gene lengths	3508.7 ± 4065.58; 3253.6 ± 3362.65	0.069	1	2891.8 ± 3763.76; 2523.2 ± 2914.29	0.113	1
Exon lengths	218.5 ± 235.79; 205.5 ± 225.23	0.057	1	227.9 ± 257.28; 225.9 ± 277.16	0.007	1
Intron lengths	306.4 ± 941.55; 299 ± 685.65	0.009	1	282.2 ± 914.22; 277.7 ± 653.87	0.006	1
Exon number	7.3 ± 4.73; 7 ± 4.51	0.049	1	6.2 ± 4.38; 5.6 ± 4.1	0.157	0.1303061

We used a linear mixed-effects model (lmer), with the reproduction system as a fixed effect and species as a random effect. The values in the columns represent *P*-values associated with the reproduction system, after the Bonferroni correction.

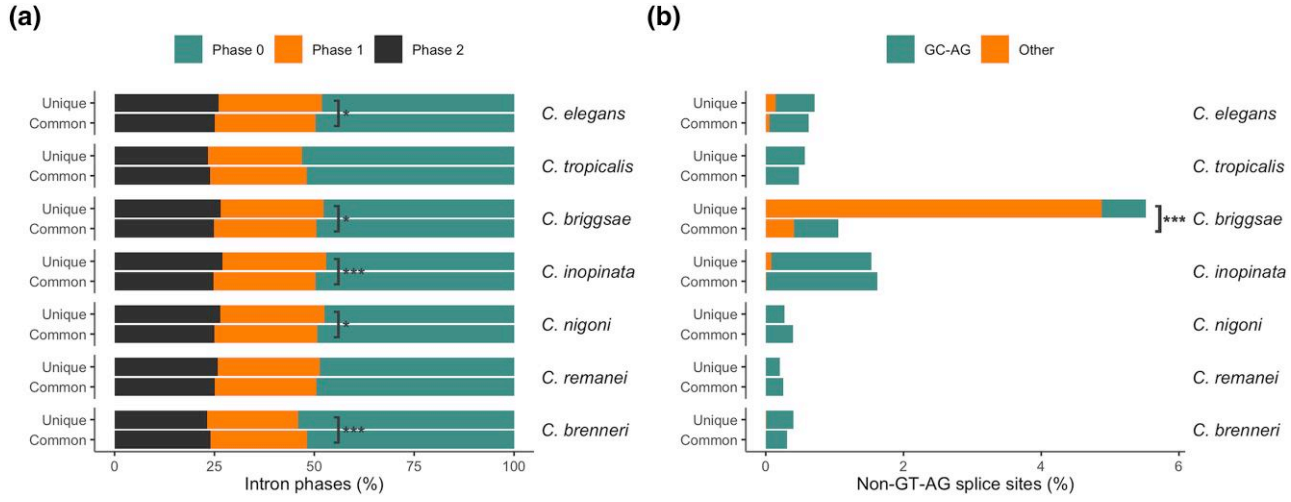

Fig. 8. Intron features in species-specific (unique) introns and introns shared with other species (common). a) Phases of introns. Colors represent the relative proportion of phase 0 introns (teal), which means that the intron is located between codons, phase 1 (orange), and phase 2 (black). Asterisks indicate the statistical significance of differences in proportions of phase 0 and nonphase 0 (combined phase 1 and phase 2) introns between unique and common introns, tested using Fisher's test after Bonferroni correction. * represents adjusted *P*-value <0.1; *** represents adjusted *P*-value <0.0001. b) Relative percentage of splice sites of introns. Only splice sites other than the most common and canonical GT-AG are shown. GC-AG sites are depicted in teal, and all other rare types are shown in orange. The asterisk denotes the statistical significance of the difference in the fraction of sites other than GT-AG between unique and shared introns, with *P*-values from Fisher's test with Bonferroni correction. *** represents that the adjusted *P*-value is <0.0001.

species, as they share just 1% of introns in single-copy orthologs. Moreover, there are no significant differences in distributions of intron number or length in species with distinct population sizes or reproduction modes, suggesting underlying functional constraints in their evolutionary dynamics.

Chromosome-level Assembly of *C. brenneri*

The new complete genome of *C. brenneri* generated here exhibits much greater continuity, featuring complete chromosomes, and scant redundancy compared with the previously available genome of *C. brenneri* (generated by the *C. brenneri* sequencing and analysis consortium), which was fragmented into thousands of scaffolds and highly duplicated due to the high residual heterozygosity within the sequenced strain. Given the extreme level of diversity observed within this species, the development of the highly inbred strain (CFB2252) was essential in achieving the desired quality and completeness of the assembled genome given the high level of diversity in this species. In addition to using long and short reads for assembly and Hi-C data for scaffolding, we also used high-accuracy long-read data from a single nematode. This approach suggests that single-individual long-read sequencing can be adapted to explore population-level variation in *C. brenneri* (and

other species with substantial genetic diversity) to overcome challenges with mapping short-read data within populations; a similar strategy was successfully employed on *Heligmosomoides* nematodes (Stevens et al. 2023). The completeness of the *C. brenneri* genome was necessary for conducting comprehensive comparative analyses across nematode species, such as macro- and micro-synteny, as well as genome and gene organization evaluations.

Conservative Macrosynteny and Domain-like Organization of the Genome in *C. brenneri*

Caenorhabditis brenneri, like other rhabditid nematodes, displays notable stability of karyotype, five autosomes, and one sex chromosome (Nigon 1949; Brenner 1974; De La Rosa et al. 2021). Syntenic reconstruction across the seven *Caenorhabditis* species with fully assembled genomes supports previous observations of multiple intrachromosomal inversions and rearrangements characterized by minimal interchromosomal changes but frequent intrachromosomal translocations and inversions (Stein et al. 2003; Hillier et al. 2007; Kanzaki et al. 2018; Stevens et al. 2020; Teterina et al. 2020; Sun et al. 2022). This conservative macrosynteny, characterized by fusion and intrachromosomal rearrangement of seven ancestral chromosome elements named Nigon elements, has been previously described throughout the Rhabditid nematode phylogeny (Tandonnet et al. 2019; De La Rosa et al. 2021). Currently, it has been proposed that the stability of nematode karyotype and such macrosynteny can be attributed to some combination of the control of chromosome pairing during meiosis and other mechanisms that regulate meiotic crossovers and genome integrity (McKim et al. 1993; Villeneuve 1994; Hillers and Villeneuve 2003; Lui and Colaiácovo 2013; Bhargava et al. 2020; Dokshin et al. 2020) and that potentially direct selection against interchromosomal translocation (Hillier et al. 2007).

Like other *Caenorhabditis* species, *C. brenneri* exhibits a domain-like organization for each chromosome with higher gene and lower repeat density in the central domains (The *C. elegans* Sequencing Consortium 1998; Parkinson et al. 2004; Cutter et al. 2009; Stevens et al. 2020; Woodruff and Teterina 2020; Carlton et al. 2022). Similarly, we also found greater conservation in the central regions of chromosomes, likely caused by similar patterns of suppression of recombination and linked selection observed in other *Caenorhabditis* species (Rockman and Kruglyak 2009; Ross et al. 2011; Teterina et al. 2023). Moreover, the central chromosomal domains have higher activity and more euchromatin compared with more inactive and heterochromatic arms in some *Caenorhabditis* species (Garrigues et al. 2015; Cabianca et al. 2019; Teterina et al. 2020; De La Cruz-Ruiz et al. 2023), potentially contributing to their greater level of conservation.

Decay in Microsynteny and Lack of Conservation in Gene Organization

Despite conservation of macrosynteny—most orthologous genes located on the same chromosome across species—nematodes display a striking lack of microsynteny (Coghlan and Wolfe 2002; Stein et al. 2003; Hillier et al. 2007; Kanzaki et al. 2018; Stevens et al. 2020; Teterina et al. 2020; Sun et al. 2022), with frequent rearrangements at the level of syntenic blocks and variable intron/exon structure at the level of individual genes. For example, we found that approximately half of the syntenic blocks are in reverse orientation, with the reversed blocks generally on average shorter than those in the forward orientation. Such structural changes in the genome are accompanied by a remarkably high rate of gene birth and death (Adams et al. 2023; Ma et al. 2024), resulting in low correlations in orthogroup sizes across the species studied here. Many gene families in nematodes are organized in clusters of co-located homologous genes, with some genes even being expressed together in an operon-like manner (Roy et al. 2002; Chen and Stein 2006; Thomas 2006; Thomas and Robertson 2008; Cutter et al. 2009; Reinke and Cutter 2009; Pettitt et al. 2014). Rapid gene turnover in nematodes is likely driven by an extreme rate of duplication (Woollard 2005; Lipinski et al. 2011; Farslow et al. 2015; Konrad et al. 2018), and a high rate of gene loss (Konrad et al. 2018), which makes it several times higher than the per-nucleotide substitution mutation rate (Denver et al. 2009; Denver et al. 2012; Konrad et al. 2019; Saxena et al. 2019). However, there is no clear relationship between rates of gene turnover and population size across species. Given the high rate of divergence in gene composition among species, analysis of variation within-species is likely to be helpful to deepen our understanding of these processes.

At the level of individual genes, nematodes are known to lose introns at a high rate (Kent and Zahler 2000; Stein et al. 2003; Cho et al. 2004; Coghlan and Wolfe 2004; Kiontke et al. 2004; Logsdon 2004; Roy and Gilbert 2005; Roy and Penny 2006; Ma et al. 2022). In particular, they tend to mostly lose short introns, usually precisely, more frequently on the 3′-end of the gene, and preferentially in genes with high expression (Ma et al. 2022). Indeed, as has been previously observed for nematode genomes, we find very low conservation in exon/intron structures of 1-to-1 orthologues, with only 1% of common introns in seven species with complete genomes compared here (Cho et al. 2004; Coghlan and Wolfe 2004). In contrast, other species exhibit high conservation of intron positions, with 99.95% conserved introns in the comparison of 1,576 gene pairs between humans and mice (Roy et al. 2003), and 93.3% common introns among 11 *Drosophila* species (Coulombe-Huntington and Majewski 2007). An apparent

exception to this pattern is the somewhat conserved pattern of intron structure for extremely conserved single-copy orthologs that show very slow rates of evolution across nematodes as a whole (Ma et al. 2022). One mechanism that could explain the dynamic nature of exon/intron evolution could potentially be the high level of gene birth and death in nematodes, which would allow new introns to arise and disappear in paralogs in a somewhat cryptic fashion. This pattern has been described for a subset of gene families within these species (Robertson 1998; Cho et al. 2004). The answer to the balance between macro conservation and micro genomic churn is likely to be better resolved as we build more high-quality assemblies at a variety of phylogenetic scales that allow micro-syntentic relationships to be assessed with adequate confidence.

Overall, our data reveals an apparent conundrum: introns apparently turn over in location in a very dynamic fashion within genes, yet our major observation is that the distribution of intron number and length across the entire genome is very similar for each species. How can both be true? The answer may lie in the functional role that introns play: mRNA processing and expression. Specifically, splicing of introns plays a critical role in facilitating the export of mRNA from the nucleus to the cytoplasm (Shiimori et al. 2013; Katahira 2015; Xie and Ren 2019; Zheleva et al. 2019). Most importantly for this discussion, introns regulate gene expression in nematodes, especially when they are located on 5′-end of the gene (Bieberstein et al. 2012; Crane et al. 2019; Sands et al. 2021), with the insertion of even a single intron increasing gene expression by as much as 50% (Crane et al. 2019). Moreover, careful experimentation in *C. elegans* molecular biology has shown that it is best to use at least three introns when creating artificial transgenes in order to achieve an optimal level of expression (Fire 1995). Interestingly, this is close to the average number of introns in all species (Fig. 7b). Furthermore, introns can regulate expression of genes through the alteration of chromatin marks (Bieberstein et al. 2012; Jo and Choi 2019), as well as generating biased allele-specific expression (Sands et al. 2021).

The efficiency of intron splicing depends on the nature of the splice sites themselves, and nematodes have U2-type of spliceosomes that only require a short motif for site recognition (Wahl et al. 2009; Bartschat and Samuelsson 2010). Roughly half of introns across all studied species were in phase 0, such that they do not disrupt a codon. This bias, typical for U2 introns, could be associated with nonrandom intron insertion, codon bias, or selection to preserve codon integrity, in keeping with studies in other nonnematode species (Fedorov et al. 1992; Long et al. 1995; Nguyen et al. 2006; Rogozin et al. 2012). Overall, the global hypothesis most consistent with the pattern of intron structure across the broad diversity of these species is that introns themselves are functionally important and clearly under selection (no significant signal from population size on intron evolution), but it is the presence of introns that is important and not necessarily their precise placement within a gene.

Genome Organization in Selfing and Outcrossing Species

An early observation regarding genomic evolution in *Caenorhabditis* was that the genomes of selfing nematodes tend to be substantially smaller than genomes of their outcrossing relatives. This 10% to 30% reduction in size turns out to be mostly attributable to extensive loss of genes and flanking regions (Bird et al. 2005; Fierst et al. 2015; Yin et al. 2018; Stevens et al. 2019; Adams et al. 2023). Reduction of some regions of the genome can in fact be driven by adaptation to self-fertilization itself rather than an epiphenomenon of the reduction in effective population size under selfing. For instance, in *C. briggsae* and other selfing species, some genes with male-biased expression, such as sperm competition proteins, seem to have been preferentially lost (Yin et al. 2018). Furthermore, hybrids of *C. briggsae* and *C. nigoni* show sex-biased gene misregulation, especially in spermatogenesis (Sánchez-Ramírez et al. 2021). However, another explanation for the difference in genome sizes between selfing and outcrossing species that has been suggested is the expansion and duplication of coding genes in the genomes of outcrossing species (Kanzaki et al. 2018; Stevens et al. 2019; Adams et al. 2023). However, we found only minor changes in the size of shared orthogroups, with outcrossing species tending to have more species-specific orthogroups.

A final explanation for the differences in genome sizes arises from an interaction of genome-level mutational processes and population genetics. In particular, it has been demonstrated that deletions are more common than insertions in *C. elegans* and that these deletions are predominantly located in intergenic regions as well as being more frequently found on chromosome arms (Konrad et al. 2019; Hwang and Wang 2021; Zhang et al. 2022). Reduced population size and a very low effective recombination rate within selfing species mean that deletions have a higher probability of fixing than in outcrossing species (Thomas et al. 2012b; Fierst et al. 2015; Hartfield 2016; Hartfield et al. 2017). With our new genomic data, we also find that the main distinction between selfing and outcrossing species is in the fraction of intergenic DNA and number of exons/genes (Fierst et al. 2015; Stevens et al. 2019; Adams et al. 2023). While all of these processes are likely to be involved in driving genomic evolution across these species, we need to move toward a full population genomics of independently assembled genomes segregating within natural populations to tease them apart any

more fully than the current state of descriptive comparison among species.

Testing the Hypothesis that Effective Population Size Structures the Evolution of Genomic Organization

The species included here range from those with little to no polymorphism within populations (*C. tropicalis*) to polymorphism at the level of one SNP every 10 bp (*C. brenneri*). Surely if population size is the main driver of the evolution of many genomic features, then we should see that signature here (Lynch 2002; Lynch and Richardson 2002; Lynch and Conery 2003; Kiontke et al. 2004). Despite this expectation, we find that the distributions of gene, exon, intron lengths, and exon number are mostly similar across all species, and no significant differences are found either between selfing and outcrossing species or among outcrossing species with orders of magnitude differences in effective population sizes. One possible explanation is that selfing species may not have had sufficient evolutionary time under reduced Ne for these genomic features to diverge significantly, as selfing in *C. elegans* and *C. briggsae* emerged relatively recently, ~4 million years ago (Cutter et al. 2008). However, simulations suggest that, given their population sizes, that timeframe should have been enough for selection to create detectable differences from their outcrossing ancestors (Teterina et al. 2023). The observation of frequent intron loss and the stability of the number of introns across species with very different population sizes suggests instead that natural selection on the balance between loss and functional necessity of introns best explains these patterns. And while the influence of population size on the evolution of putatively "nearly neutral" genetic elements has certainly been an important feature of many theories of molecular evolution, it has also been simultaneously suggested that once these features come to be present within genomes, they will quickly evolve other functional roles (Lynch 2007; Jo and Choi 2015; Girardini et al. 2023). The molecular biology of *Caenorhabditis* nematodes, including truly hyper-polymorphic species, suggests that the latter circumstance dominates the former and that most features of the genome are under direct selection, even if it is not selection on a specific sequence of DNA.

Materials and Methods

Strain Generation and Sequencing

The *C. brenneri* line CFB2252 was initially isolated as strain VX0044, a single gravid female from a rotting marigold flower in Shyamnagar, India by the Cutter Lab (University of Toronto, Dey et al. 2013), who subsequently inbred it though sib-mating under standard nematode lab-rearing conditions (Brenner 1974) for 20 generations, with population expansion every five generations to help purge the accumulation of large-effect deleterious mutations. The endpoint of this inbreeding resulted in strain VX0225. This strain was obtained from the Cutter Lab by the Baer Lab, who performed another 22 generations of single-female descent before expanding the population to large size and maintained it by chunking at 3-day intervals for eight generations. VX0225 was derived from that population by creating a single line maintained by single-female descent for an additional 65 generations. That line was re-labeled as CFB2252 and frozen. Thus, the final sequencing line, CFB2252 was inbred for roughly 100 generations from the initial collection.

For PacBio long-read sequencing of the *C. brenneri* CFB2252 strain, genomic DNA was isolated from a starved L1 larvae population (~3,000 worms) following Proteinase K digestion and purification with the Genomic DNA Clean and Concentrator-25 kit (ZYMO). For PacBio IsoSeq (transcript isoform) sequencing, we isolated total RNA from a mixed-stage population with multiple freeze/crack steps in liquid nitrogen before utilizing the Zymo DirectZol RNA microprep kit (R2062). For individual worm long-read sequencing, L4 females were picked to a separate plate to avoid mating. Twenty-four hours later at 20 °C, the adult virgin female worms were picked to a watch glass with M9 taking care not to transfer bacteria from the plate. After 10 min, an individual female nematode was transferred to a fresh watch glass with M9 and then to a microcentrifuge tube with 10 μL of water and Proteinase K (2 mg/mL final). Following incubation at 58 °C for 1 h, the genomic DNA was isolated via 1× AMPureXP bead purification. The eluted genomic DNA sample was amplified using the PacBio Ultra-Low DNA Input Workflow. PacBio CLR sequencing with the Sequel I platform was used for all long-read sequencing. For short read sequencing of genomic DNA, we used the Nextera XT DNA Library Preparation Kit to generate Illumina compatible libraries. Similarly, Hi-C sequencing using the Phase Genomics platform utilized genomic DNA prepared above. For mRNA sequencing, libraries were generated with the Kapa mRNA Hyper prep kit (KK8580) starting with 100 ng of total RNA. The Illumina HiSeq 4000 platform was used for all short-read sequencing. All sequencing was completed at the University of Oregon Genomics and Cell Characterization Core facility (GC3F) (https://gc3f.uoregon.edu/).

Genome Assembly

Potential adapter sequences were removed from each long read using removesmartbell.sh from bbmap v.38.16 (Bushnell 2017). Hi-C read and other short reads were trimmed and filtered using skewer v.0.2.2 (Jiang et al. 2014). Genome assembly was conducted with PacBio CLR reads by FALCON v.1.2.3 and FALCON-Unzip v.1.1.3

(Chin et al. 2016), followed by additional phasing with Hi-C reads by FALCON-Phase v.1.0.0 (Kronenberg et al. 2021). Contigs were polished with Illumina short reads by Pilon v.1.23 (Walker et al. 2014) with bwa v.0.7.17 (Li 2013), and samtools v.1.5 (Li et al. 2009). Then we performed contig scaffolding with Hi-C data following the 3D-DNA pipeline (Dudchenko et al. 2017) with run-asm-pipeline.sh v.180922 and Juicer v.1.6.2 (Durand et al. 2016), BWA and samtools. Then scaffolds were polished with HiFi reads by RACON v1.4.20 (Vaser et al. 2017). HiFi reads were corrected using Illumina read with fmlrc v1.0.0 (Wang et al. 2018). In addition, we assembled HiFi reads from an individual nematode of the reference strain (CFB2252) using hifiasm v.0.15.1-r334 (Cheng et al. 2021), and assembled short read obtained from the pool of individuals by SGA v.0.10.15 (Simpson and Durbin 2012). Contigs from those two assemblies were used for validation of the final genome assembly. Gaps in the *C. brenneri* genome were filled using corrected and uncorrected HiFi reads by LR_GapCloser (Xu et al. 2019) and TGS_GapCloser (Xu et al. 2020). We assembled and annotated a mitochondrial genome from HiFi reads with MitoHiFi v3.0.0 (Uliano-Silva et al. 2023) using the mitochondrial genome of *C. brenneri* as a reference (the accession number in Gene Bank is NC_035244.1). We removed bacterial contamination using bloobtools v1.1.1 (Laetsch and Blaxter 2017), and validated scaffolds using HiFi reads with Inspector v1.0.1 (Chen et al. 2021). The quality of the final genome assembly was accessed with Inspector, MERQURY v2020-01-29 (Rhie et al. 2020), BUSCO v5.4.2 (Simão et al. 2015) with databases nematode_odb10 and metazoa_odb10, and coverage from HiFi, Illumina reads, hifiasm, and sga assemblies using pbmm2 from pb-assembly v1 (https://github.com/PacificBiosciences/pb-assembly), Samtools, and Bedtools v2.25.0 (Quinlan and Hall 2010). The heterozygosity was estimated for the HiFi long reads from the individual nematode from the CFB2252 strain with jellyfish v2.2.10 (Marçais and Kingsford 2011) and GenomeScope v2 (Vurture et al. 2017). We used minimap2 v2.20 (Li 2018) and bedtools to estimate genome coverage of individual nematode HiFi data in 100 kb genomic windows. Additionally, we estimated the fraction of residual variation, as both heterozygous and homozygous alternative alleles, across the genome using Picard v2.17.6 (Broad Institute 2018), GATK v4.1.4.1 (McKenna et al. 2010), samtools, and bedtools. To get coordinates of repetitive regions, we used generate_masked_ranges.py (Cook 2014) on the hard-masked version of the genome and excluded these repeats from the genomic windows.

Genome Annotation and Repeat Masking

We masked repetitive sequences in *C. brenneri* genome with an approach described in detail in Teterina et al. (2020), with RepeatMasker v.4.0.7 (https://github.com/rmhubley/RepeatMasker), RepeatModeler v.1.0.11 (http://www.repeatmasker.org/RepeatModeler/), detectMITE v.2017 to 2004–2025 (Ye et al. 2016), transposonPSI (http://transposonpsi.sourceforge.net), LTRharvest and LTRdigest from GenomeTools v.1.5.11 (Gremme et al. 2013), tRNAscan v.1.3.1 (Lowe and Eddy 1997), Blast.2.2.29+ (Altschul et al. 1990), and USEARCH v.8.0 (Edgar 2010). We annotated the genome using RNA-seq data and single-molecule long-read RNA sequencing (Iso-Seq), and protein sequences of *C. briggsae* (PRJNA10731), *C. brenneri* (PRJNA20035), *C. elegans* (PRJNA13758), and *C. remanei* (PRJNA577507) from WormBase 287 and *C. wallacei* (strain JU1898) and *C. doughertyi* (strain JU177) from the Caenorhabditis Genomes Project v.2 (http://download.caenorhabditis.org/v2/).

We trimmed RNA-seq reads by skewer and mapped them to the *C. brenneri* genome with STAR v2.5.3 (Dobin et al. 2013); these alignments were used as evidence for annotation with BRAKER v2.1.0 (Hoff et al. 2019). Then, we used *Caenorhabditis* proteins and BRAKER gene predictions to perform annotation with MAKER. v.2.31.9 (Cantarel et al. 2008). The Iso-Seq data was processed using the isoseq3 pipeline (https://github.com/ylipacbio/IsoSeq3). Then annotation from Braker, Maker and collapsed isoforms from isoseq3 pipeline were combined by Evidence Modeler v.1.1.1 (Haas et al. 2008). Then we leveraged splicing prediction using RNA-seq alignments with the PASA pipeline v2.5.3 (Haas 2003). We validated the final assembly with RNA-seq reads, blast to nematode proteins, and functional annotation with Interproscan v. 5.27-66.0 (Jones et al. 2014), and BUSCO. Several genes were filtered as their exons were located in the masked regions of the genome and classified as repetitive elements by RepeatClassifier from RepeatModeller. We updated gene features and formatted annotation files using AGAT tools v1.0.0 (https://github.com/NBISweden/AGAT) and custom scrips provided in the GitHub repository for this project.

Comparative Genomic Analysis

Divergence and Conservation

We aligned the new soft-masked genome of *C. brenneri* with genomes of *C. elegans* (PRJNA13758), *C. remanei* (PRJNA577507), and *C. tropicalis* (PRJNA53597) from WormBase WS287, and *C. wallacei*, *C. doughertyi*, *C. sp44*, *C. sp48*, *C. sp51*, and *C. sp54* from the Caenorhabditis Genomes Project v.2 using progressiveCactus (Armstrong et al. 2020). The alignments were split into 50 kb windows by hal2maf_split.pl from HAL tools (Hickey et al. 2013), converted to MSA format with msaconverter (https://github.com/linzhi2013/msaconverter) and filtered by maffilter (Dutheil et al. 2014). The maximum-likelihood trees were constructed using IQ-TREE v2.2.0.3 (Nguyen et al. 2015). We got

topologies of the trees, a total length of *C. brenneri* and *C. sp48* clade, or solely the *C. brenneri* branch from a few trees where they were not clustered together. The total tree heights were estimated using package ape v5.3 (Paradis et al. 2004) in R v3.5.0 (R Core Team 2018).

We conducted a conservation sequence analysis using progressive Cactus alignments, which were converted from HAL to MAF format using hal2maf from the HAL tools, and utilized phyloFit and phastCons tools from Phast v1.5 (Siepel et al. 2005). First, we generated an initial model with phyloFit using the following tree "(((((*C. brenneri*, *C. sp48*), (*C. sp44*, *C. sp51*)), (*C. doughertyi*, (*C. tropicalis*, *C. wallacei*))), *C. remanei*), *C. elegans*, *C. sp54*)" and the whole genome alignments. Following that, we retrained the model for each chromosome with phastCons, and finally applied these models to estimate conservation in 50 kb window alignments for the corresponding chromosomes and calculated the average scores within those windows. The visualization was done in R with packages ape, ggplot2 v3.4.2 (Wickham 2009), ggtree v1.14.6 (Yu et al. 2017), and patchwork v1.1.1 (Pedersen 2024).

Synteny, Orthology, and Gene Structures

For orthology analysis, we used six *Caenorhabditis* species with chromosome-scale genome assemblies: genomes and canonical annotations of *C. elegans* (PRJNA13758), *C. remanei* (PRJNA577507), *C. inopinata* (PRJDB5687), *C. briggsae* (PRJNA10731), *C. nigoni* (PRJNA384657) were obtained from the WormBase WBPS18 and the genome and annotation for NIC58 strain of *C. tropicalis* were sourced from (Noble et al. 2021). The corresponding protein sequences were generated from the longest transcripts by AGAT tools. Orthology analysis was conducted by Orthofinder v2.5.5 (Emms and Kelly 2019). Then we estimated the number of exons, lengths of genes, exons, and introns without all genes that have orthogroups and genes that are in single-copy in all seven species (1-to-1) orthogroups. We used R packages stats v3.5.0, dplyr v1.0.7 (Wickham et al. 2022), tidyr v1.1.3 (Wickham 2021), esvis v0.3.1 (Anderson 2020), data.table v1.12.8 (Dowle et al. 2021), reshape2 v1.4.4 (Wickham 2007), ggplot2, corrplot v0.90 (Wei and Simko 2021), RColorBrewer v1.1.2 (Neuwirth 2014), ggplotify v0.0.8 (Yu 2021), gridGraphics v0.5.1 (Murrell and Wen 2020), patchcwork to compare and visualize these distributions and difference between species. The specific tests are indicated in figure legends and the main text.

The inference of syntenic blocks was performed based on the results from Orthofinder2 by R package GENESPACE v1.2.3 (Lovell et al. 2022). We visualized the riparian plot with *C. brenneri* as the reference genome, and estimated the number of genes, exons, introns, and fractions of exons/introns within corresponding syntenic blocks. As some species were lacking the annotation of the 3′ and 5′ UTR and other, we called regions that were neither exons nor introns as "intergenic." We then calculated the ratios of these measurements for species pairs in relation to *C. brenneri*. This allowed us to explore the connection between genome organization and the differences in population size ratios among species, which sometimes span orders of magnitude. However, due to a lack of population size estimates for certain species like *C. nigoni*, we focused on the classification of selfers versus outcrossers using ratios of the number of exons, introns, and the length of exons, introns, and intergenic regions. We replicated syntenic blocks proportionally to their sizes in the *C. brenneri* genome and trained a GLM and a gradient boosting model to determine whether the ratios correspond to selfing species (*C. elegans* and *C. tropicalis*) versus outcrossing *C. brenneri* or outcrossing species (*C. remanei* and *C. inopinata*) versus *C. brenneri*. Then we evaluated the model with selfing *C. briggsae* versus *C. brenneri* and outcrossing *C. nigoni* versus *C. brenneri*. For the GLM and gradient boosting model (XGboost) approaches, we utilized R packages glmnet v2.0.18 (Friedman et al. 2010), xgboost v1.6.0.1 (Chen et al. 2022), tidyr, dplyr, caret v6.0.8 (Kuhn 2020), broom v0.7.9 (Robinson et al. 2021), pROC v1.18.0 (Robin et al. 2011), caTools v1.17.1.2 (Tuszynski 2019), gtable v0.3.0 (Wickham and Pedersen 2019), DALEX v2.4.3 (Biecek 2018).

Comparison of Exon/Intron Gene Structures

We explored the exon/intron structures of single-copy orthologs determined by OrthoFinder2 and corresponding gene annotations using GenePainter v2.0 (Hammesfahr et al. 2013; Mühlhausen et al. 2015). Then we estimated the sizes of the introns, accessed phases of the CDS (phase 0,1, or 2) that follow the intron, and obtained splice sites using *agat_sp_extract_sequences.pl* form AGAT tools v1.2.0 to get splice sites. Then we parsed sizes, phases, and splice sites of introns with gene exon/intron structures alignments from GenePainter and compared unique and common introns as well as estimated association and difference of introns shared in pairs of nematode species using R packages dplyr, reshape2, lme4 v1.1.27.1 (Bates et al. 2015), lmerTest v3.1.3 (Kuznetsova et al. 2017), MASS v.7.3.54 (Venables and Ripley 2002), ggplot2, patchwork, ggplotify, gridGraphics, corrplot, and phylotools.

Supplementary Material

Supplementary figs. (S1 to S10) and supplementary table S1, Supplementary Material online are available at *Genome Biology and Evolution* online (Supplementary material).

Acknowledgments

We thank Mark Blaxter and the *Caenorhabditis* Genomes team for permission to use the unpublished genome sequences for a number of species for some of our analyses. Asher Cutter kindly provided the semi-inbred progenitor of the CFB2252 line. We thank members of the Phillips laboratory and Kern-Ralph co-lab at the University of Oregon for the helpful discussions and the reviewers for their helpful suggestions. We also thank the University of Oregon Genomics and Cell Characterization Core Facility (GC3F) for support, and the Research Advanced Computing Services team at the University of Oregon for assistance with the high-performance computer Talapas.

Funding

The project was supported by funding from the National Institutes of Health to P.C.P. (R35GM131838).

Data Availability

Genomic and transcriptomic data generated in this work available at the European Nucleotide Archive database (ENA, https://www.ebi.ac.uk/ena) under BioProject ID PRJEB72296, which includes eleven experiments (ERX12369330-ERX12369340). The *C. brenneri's* genome accession number is GCA_964036135.1. Copies of the current version of the genome and annotation of *C. brenneri* used in the analysis are available at Figshare under doi: https://doi.org/10.6084/m9.figshare.25988629, all files necessary to estimate statistics and generate the figures available at Figshare under doi:https://doi.org/10.6084/m9.figshare.25988626.v2. Scripts for the analysis and visualization available at https://github.com/phillips-lab/Cbren_genome.

Literature Cited

Adams PE, Eggers VK, Millwood JD, Sutton JM, Pienaar J, Fierst JL. Genome size changes by duplication, divergence, and insertion in *Caenorhabditis* worms. Mol Biol Evol. 2023:40(3):msad039. https://doi.org/10.1093/molbev/msad039.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderson D. 2020. esvis: visualization and estimation of effect sizes. Available from: https://CRAN.R-project.org/package=esvis

Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature. 2020:587(7833):246–251. https://doi.org/10.1038/s41586-020-2871-y.

Barrière A, Félix M-A. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. Curr Biol. 2005:15(13):1176–1184. https://doi.org/10.1016/j.cub.2005.06.022.

Barrière A, Yang S-P, Pekarek E, Thomas CG, Haag ES, Ruvinsky I. Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. Genome Res. 2009:19(3):470–480. https://doi.org/10.1101/gr.081851.108.

Bartschat S, Samuelsson T. U12 type introns were lost at multiple occasions during evolution. BMC Genomics. 2010:11(1):106. https://doi.org/10.1186/1471-2164-11-106.

Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015:67(1):1–48. https://doi.org/10.18637/jss.v067.i01.

Bhargava V, Goldstein CD, Russell L, Xu L, Ahmed M, Li W, Casey A, Servage K, Kollipara R, Picciarelli Z, et al. GCNA preserves genome integrity and fertility across species. Dev Cell. 2020:52(1):38–-52.e10. https://doi.org/10.1016/j.devcel.2019.11.007.

Bieberstein NI, Carrillo Oesterreich F, Straube K, Neugebauer KM. First exon length controls active chromatin signatures and transcription. Cell Rep. 2012:2(1):62–68. https://doi.org/10.1016/j.celrep.2012.05.019.

Biecek P. DALEX: explainers for complex predictive models in R. J Mach Learn Res. 2018:19:1–5. https://jmlr.org/papers/v19/18-416.html.

Bird DM, Blaxter ML, McCarter JP, Mitreva M, Sternberg PW, Thomas WK. A white paper on nematode comparative genomics. J Nematol. 2005:37:408–416.

Brenner S. The genetics of *Caenorhabditis elegans*. Genetics. 1974:77(1):71–94. https://doi.org/10.1093/genetics/77.1.71.

Broad Institute. Picard tools. 2018 [accessed 2024 Jul 1]. http://broadinstitute.github.io/picard/.

Burns KH. Repetitive DNA in disease. Science. 2022:376(6591):353–354. https://doi.org/10.1126/science.abl7399.

Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. 2014 [accessed July 1, 2024]. Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). https://www.osti.gov/biblio/1241166

Cabianca DS, Muñoz-Jiménez C, Kalck V, Gaidatzis D, Padeken J, Seeber A, Askjaer P, Gasser SM. Active chromatin marks drive spatial sequestration of heterochromatin in *C. elegans* nuclei. Nature. 2019:569(7758):734–739. https://doi.org/10.1038/s41586-019-1243-y.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008:18(1):188–196. https://doi.org/10.1101/gr.6743907.

Carlton PM, Davis RE, Ahmed S. Nematode chromosomes. Genetics. 2022:221(1):iyac014. https://doi.org/10.1093/genetics/iyac014.

Carmel L, Chorev M. The function of introns. Front Genet. 2012:3:55. https://doi.org/10.3389/fgene.2012.00055.

Charlesworth B, Barton N. Genome size: does bigger mean worse? Curr Biol. 2004:14(6):R233–R235. https://doi.org/10.1016/j.cub.2004.02.054.

Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations on neutral molecular variation. Genetics. 1993:134(4):1289–1303. https://doi.org/10.1093/genetics/134.4.1289.

Chen N, Stein LD. Conservation and functional significance of gene topology in the genome of *Caenorhabditis elegans*. Genome Res. 2006:16(5):606–617. https://doi.org/10.1101/gr.4515306.

Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, et al. xgboost: extreme gradient boosting. 2022 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=xgboost.

Chen Y, Zhang Y, Wang AY, Gao M, Chong Z. Accurate long-read de novo assembly evaluation with inspector. Genome Biol. 2021:22(1):312. https://doi.org/10.1186/s13059-021-02527-4.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016:13(12):1050–1054. https://doi.org/10.1038/nmeth.4035.

Cho S, Jin S-W, Cohen A, Ellis RE. A phylogeny of *Caenorhabditis* reveals frequent loss of introns during nematode evolution. Genome Res. 2004:14(7):1207–1220. https://doi.org/10.1101/gr.2639304.

Coghlan A, Wolfe KH. Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. Genome Res. 2002:12(6):857–867. https://doi.org/10.1101/gr.172702.

Coghlan A, Wolfe KH. Origins of recently gained introns in *Caenorhabditis*. Proc Natl Acad Sci U S A. 2004:101(31):11362–-11367. https://doi.org/10.1073/pnas.0308192101.

Cook DE. Generate the masked ranges within a fasta file. *Gist*. 2014 [accessed 2024 Jul 1]. https://gist.github.com/danielecook/cfaa5c359d99bcad3200.

Coop G, Ralph P. Patterns of neutral diversity under general models of selective sweeps. Genetics. 2012:192(1):205–224. https://doi.org/10.1534/genetics.112.141861.

Coulombe-Huntington J, Majewski J. Characterization of intron loss events in mammals. Genome Res. 2007:17(1):23–32. https://doi.org/10.1101/gr.5703406.

Crane MM, Sands B, Battaglia C, Johnson B, Yun S, Kaeberlein M, Brent R, Mendenhall A. In vivo measurements reveal a single 5′-intron is sufficient to increase protein expression level in *Caenorhabditis elegans*. Sci Rep. 2019:9(1):9192. https://doi.org/10.1038/s41598-019-45517-0.

Cutter AD. Nucleotide polymorphism and linkage disequilibrium in wild populations of the partial selfer *Caenorhabditis elegans*. Genetics. 2006:172(1):171–184. https://doi.org/10.1534/genetics.105.048207.

Cutter AD. Reproductive transitions in plants and animals: selfing syndrome, sexual selection and speciation. New Phytol. 2019:224(3):1080–1094. https://doi.org/10.1111/nph.16075.

Cutter AD, Baird SE, Charlesworth D. High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics. 2006:174(2):901–913. https://doi.org/10.1534/genetics.106.061879.

Cutter AD, Dey A, Murray RL. Evolution of the *Caenorhabditis elegans* genome. Mol Biol Evol. 2009:26(6):1199–1234. https://doi.org/10.1093/molbev/msp048

Cutter AD, Morran LT, Phillips PC. *Males, outcrossing, and sexual selection in Caenorhabditis* nematodes. Genetics. 2019:213(1):27–57. https://doi.org/10.1534/genetics.119.300244.

Cutter AD, Wasmuth JD, Washington NL. Patterns of molecular evolution in *Caenorhabditis* preclude ancient origins of selfing. Genetics. 2008:178(4):2093–2104. https://doi.org/10.1534/genetics.107.085787.

Daubin V, Moran NA. Comment on "the origins of genome complexity.". Science. 2004:306(5698):978–978. https://doi.org/10.1126/science.1098469.

De La Cruz-Ruiz P, Rodríguez-Palero MJ, Askjaer P, Artal-Sanz M. Tissue-specific chromatin-binding patterns of *Caenorhabditis elegans* heterochromatin proteins HPL-1 and HPL-2 reveal differential roles in the regulation of gene expression. Genetics. 2023:224(3):iyad081. https://doi.org/10.1093/genetics/iyad081.

De La Rosa G, Manuel P, Thomson M, Trivedi U, Tracey A, Tandonnet S, Blaxter M. A telomere-to-telomere assembly of *Oscheius tipulae* and the evolution of rhabditid nematode chromosomes. 2021:11(1):jkaa020. https://doi.org/10.1093/g3journal/jkaa020.

Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc Natl Acad Sci U S A. 2009:106(38):16310–16314. https://doi.org/10.1073/pnas.0904895106.

Denver DR, Wilhelm LJ, Howe DK, Gafner K, Dolan PC, Baer CF. Variation in base-substitution mutation in experimental and natural lineages of *Caenorhabditis* nematodes. Genome Biol Evol. 2012:4(4):513–522. https://doi.org/10.1093/gbe/evs028.

de Souza SJ, Long M, Klein RJ, Roy S, Lin S, Gilbert W. Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc Natl Acad Sci U S A. 1998:95(9):5094–5099. https://doi.org/10.1073/pnas.95.9.5094.

Dey A, Chan CKW, Thomas CG, Cutter AD. Molecular hyperdiversity defines populations of the nematode *Caenorhabditis brenneri*. Proc Natl Acad Sci U S A. 2013:110(27):11056–11060. https://doi.org/10.1073/pnas.1303057110.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dokshin GA, Davis GM, Sawle AD, Eldridge MD, Nicholls PK, Gourley TE, Romer KA, Molesworth LW, Tatnell HR, Ozturk AR, et al. GCNA interacts with spartan and topoisomerase II to regulate genome stability. Dev Cell. 2020:52(1):53–68.e6. https://doi.org/10.1016/j.devcel.2019.11.006.

Dolgin ES, Charlesworth B, Baird SE, Cutter AD. Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. Evolution. 2007:61(6):1339–1352. https://doi.org/10.1111/j.1558-5646.2007.00118.x.

Dolgin ES, Félix M-A, Cutter AD. Hakuna Nematoda: genetic and phenotypic diversity in African isolates of *Caenorhabditis elegans* and *C. briggsae*. Heredity (Edinb). 2008:100(3):304–315. https://doi.org/10.1038/sj.hdy.6801079.

Dowle M, Srinivasan A, Gorecki J, Chirico M, Stetsenko P, Short T, Lianoglou S, Antonyan E, Bonsch M, Parsonage H, et al. data.table: Extension of "data.frame." 2021 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=data.table.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the *Aedes aegypti* genome using hi-C yields chromosome-length scaffolds. Science. 2017:356(6333):92–95. https://doi.org/10.1126/science.aal3327.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. Cell Syst. 2016:3:95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Dutheil JY, Gaillard S, Stukenbrock EH. MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics. 2014:15:53. https://doi.org/10.1186/1471-2164-15-53.

Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010:26(19):2460–2461. https://doi.org/10.1093/bioinformatics/btq461.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Farslow JC, Lipinski KJ, Packard LB, Edgley ML, Taylor J, Flibotte S, Moerman DG, Katju V, Bergthorsson U. Rapid increase in frequency of gene copy-number variants during experimental evolution in *Caenorhabditis elegans*. BMC Genomics. 2015:16:1044. https://doi.org/10.1186/s12864-015-2253-2.

Fedorov A, Suboch G, Bujakov M, Fedorova L. Analysis of nonuniformity in intron phase distribution. Nucleic Acids Res. 1992:20(10): 2553–2557. https://doi.org/10.1093/nar/20.10.2553.

Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, Ahearne TE, Cutter AD, Phillips PC. Reproductive mode and the evolution of genome size and structure in *Caenorhabditis* nematodes. PLoS Genet. 2015:11(6):e1005323. https://doi.org/10.1371/journal.pgen.1005323.

Fire A. Fire lab *C. elegans* vector kit 1995. 1995. [accessed 2024 Jul 1]. https://media.addgene.org/cms/files/Vec95.pdf.

Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw. 2010:33(1): 1–22. https://doi.org/10.18637/jss.v033.i01.

Galtier N. Half a century of controversy: the neutralist/selectionist debate in molecular evolution. Genome Biol Evol. 2024:16(2): evae003. https://doi.org/10.1093/gbe/evae003.

Garrigues JM, Sidoli S, Garcia BA, Strome S. Defining heterochromatin in *C. elegans* through genome-wide analysis of the heterochromatin protein 1 homolog HPL-2. Genome Res. 2015:25(1):76–88. https://doi.org/10.1101/gr.180489.114.

Girardini KN, Olthof AM, Kanadia RN. Introns: the "dark matter" of the eukaryotic genome. Front Genet. 2023:14:1150212. https://doi.org/10.3389/fgene.2023.1150212.

Golding GB, Strobeck C. Linkage disequilibrium in a finite population that is partially selfing. Genetics. 1980:94(3):777–789. https://doi.org/10.1093/genetics/94.3.777.

Graustein A, Gaspar JM, Walters JR, Palopoli MF. Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics. 2002:161(1):99–107. https://doi.org/10.1093/genetics/161.1.99.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG, Kapraun DF, Greilhuber J, Bennett MD. Eukaryotic genome size databases. Nucleic Acids Res. 2007:35:D332–D338. https://doi.org/10.1093/nar/gkl828.

Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. IEEE/ACM Trans Comput Biol Bioinf. 2013:10:645–656. https://doi.org/10.1109/TCBB.2013.68.

Haas BJ. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003:31(19): 5654–5666. https://doi.org/10.1093/nar/gkg770.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008:9(1):R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Hammesfahr B, Odronitz F, Mühlhausen S, Waack S, Kollmar M. GenePainter: a fast tool for aligning gene structures of eukaryotic protein families, visualizing the alignments and mapping gene structures onto protein structures. BMC Bioinformatics. 2013:14: 77. https://doi.org/10.1186/1471-2105-14-77.

Hartfield M. Evolutionary genetic consequences of facultative sex and outcrossing. J Evol Biol. 2016:29(1):5–22. https://doi.org/10.1111/jeb.12770.

Hartfield M, Bataillon T, Glémin S. The evolutionary interplay between adaptation and self-fertilization. Trends Genet. 2017:33(6): 420–431. https://doi.org/10.1016/j.tig.2017.04.002.

Hickey G, Paten B, Earl D, Zerbino D, Haussler D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. Bioinformatics. 2013:29(10):1341–1342. https://doi.org/10.1093/bioinformatics/btt128.

Hillers KJ, Villeneuve AM. Chromosome-wide control of meiotic crossing over in *C. elegans*. Curr Biol. 2003:13(18):1641–1647. https://doi.org/10.1016/j.cub.2003.08.026.

Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH. Comparison of *C. elegans* and *C. Briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 2007:5(7):e167. https://doi.org/10.1371/journal.pbio.0050167.

Hoff K, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Methods Mol Biol. 2019:1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

Hwang H-Y, Wang J. Fast genetic mapping using insertion-deletion polymorphisms in *Caenorhabditis elegans*. Sci Rep. 2021:11(1): 11017. https://doi.org/10.1038/s41598-021-90190-x.

Jiang H, Lei R, Ding S-W, Zhu S. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. BMC Bioinformatics. 2014:15:182. https://doi.org/10.1186/1471-2105-15-182.

Jo B-S, Choi SS. Introns: the functional benefits of introns in genomes. Genomics Inform. 2015:13(4):112–118. https://doi.org/10.5808/GI.2015.13.4.112.

Jo S-S, Choi SS. Analysis of the functional relevance of epigenetic chromatin marks in the first intron associated with specific gene expression patterns. Genome Biol Evol. 2019:11:786–797. https://doi.org/10.1093/gbe/evz033.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014:30(9): 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Kanzaki N, Tsai IJ, Tanaka R, Hunt VL, Liu D, Tsuyama K, Maeda Y, Namai S, Kumagai R, Tracey A, et al. Biology and genome of a newly discovered sibling species of *Caenorhabditis elegans*. Nat Commun. 2018:9(1):3216. https://doi.org/10.1038/s41467-018-05712-5.

Katahira J. Nuclear export of messenger RNA. Genes (Basel). 2015:6(2):163–184. https://doi.org/10.3390/genes6020163.

Kent WJ, Zahler AM. Conservation, regulation, synteny, and introns in a large-scale *C. Briggsae-C. elegans* genomic alignment. Genome Res. 2000:10(8):1115–1125. https://doi.org/10.1101/gr.10.8.1115.

Kiontke K, Gavin NP, Raynes Y, Roehrig C, Piano F, Fitch DHA. *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc Natl Acad Sci U S A. 2004:101: 9003–9008. https://doi.org/10.1073/pnas.0403094101.

Kiontke KC, Félix M-A, Ailion M, Rockman MV, Braendle C, Pénigault J-B, Fitch DH. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. BMC Evol Biol. 2011:11:339. https://doi.org/10.1186/1471-2148-11-339.

Konrad A, Brady MJ, Bergthorsson U, Katju V. Mutational landscape of spontaneous base substitutions and small indels in experimental *Caenorhabditis elegans* populations of differing size. Genetics. 2019:212(3):837–854. https://doi.org/10.1534/genetics.119.302054.

Konrad A, Flibotte S, Taylor J, Waterston RH, Moerman DG, Bergthorsson U, Katju V. Mutational and transcriptional landscape of spontaneous gene duplications and deletions in *Caenorhabditis elegans*. Proc Natl Acad Sci U S A 2018:115:7386–7391. https://doi.org/10.1073/pnas.1801930115.

Kronenberg ZN, Rhie A, Koren S, Concepcion GT, Peluso P, Munson KM, Porubsky D, Kuhn K, Mueller KA, Low WY, et al. Extended haplotype-phasing of long-read de novo genome assemblies using hi-C. Nat Commun. 2021:12(1):1935. https://doi.org/10.1038/s41467-020-20536-y.

Kuhn M. caret: classification and regression training. 2020 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=caret.

Kuznetsova A, Brockhoff PB, Christensen RHB. lmerTest package: tests in linear mixed effects models. J Stat Softw. 2017:82(13):1–26. https://doi.org/10.18637/jss.v082.i13.

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Res. 2017:6:1287. https://doi.org/10.12688/f1000research.12232.1.
Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. https://arxiv.org/abs/1303.3997, 2013, preprint: not peer reviewed.
Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, Bergthorsson U. High spontaneous rate of gene duplication in *Caenorhabditis elegans*. Curr Biol. 2011:21(4):306–310. https://doi.org/10.1016/j.cub.2011.01.026.
Logsdon JM. Worm genomes hold the smoking guns of intron gain. Proc Natl Acad Sci U S A. 2004:101(31):11195–11196. https://doi.org/10.1073/pnas.0404148101.
Long M, Rosenberg C, Gilbert W. Intron phase correlations and the evolution of the intron/exon structure of genes. Proc Natl Acad Sci U S A. 1995:92:12495–12499. https://doi.org/10.1073/pnas.92.26.12495.
Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife. 2022:11:e78526. https://doi.org/10.7554/eLife.78526.
Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997:25(5):955–964. https://doi.org/10.1093/nar/25.5.955.
Lui DY, Colaiácovo MP. Meiotic development in *Caenorhabditis elegans*. Adv Exp Med Biol. 2013:757:133–170. https://doi.org/10.1007/978-1-4614-4015-4_6.
Lynch M. Intron evolution as a population-genetic process. Proc Natl Acad Sci U S A. 2002:99(9):6118–6123. https://doi.org/10.1073/pnas.092595699.
Lynch M. The origins of genome architecture. Sunderland, MA: Sinauer; 2007.
Lynch M, Conery JS. The origins of genome complexity. Science. 2003:302(5649):1401–1404. https://doi.org/10.1126/science.1089370.
Lynch M, Richardson AO. The evolution of spliceosomal introns. Curr Opin Genet Dev. 2002:12(6):701–710. https://doi.org/10.1016/S0959-437X(02)00360-X.
Ma F, Lau CY, Zheng C. Young duplicate genes show developmental stage- and cell type-specific expression and function in *Caenorhabditis elegans*. Cell Genom. 2024:4(1):00467. https://doi.org/10.1016/j.xgen.2023.100467.
Ma M-Y, Xia J, Shu K-X, Niu D-K. Intron losses and gains in the nematodes. Biol Direct. 2022:17(1):13. https://doi.org/10.1186/s13062-022-00328-8.
Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k* -mers. Bioinformatics. 2011:27(6):764–770. https://doi.org/10.1093/bioinformatics/btr011.
Marino A, Debaecker G, Fiston-Lavier A-S, Haudry A, Nabholz B. Effective population size does not explain long-term variation in genome size and transposable element content in animals. Evol Biol. 2025:52(1):26–39. https://doi.org/10.1007/s11692-024-09643-6.
McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010:20(9):1297–1303. https://doi.org/10.1101/gr.107524.110.
McKim KS, Peters K, Rose AM. Two types of sites required for meiotic chromosome pairing in *Caenorhabditis elegans*. Genetics. 1993:134(3):749–768. https://doi.org/10.1093/genetics/134.3.749.
Mühlhausen S, Hellkamp M, Kollmar M. GenePainter v. 2.0 resolves the taxonomic distribution of intron positions. Bioinformatics. 2015:31(8):1302–1304. https://doi.org/10.1093/bioinformatics/btu798.
Murrell P, Wen Z. gridGraphics: redraw base graphics using "grid" graphics. 2020 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=gridGraphics.
Neuwirth E. RColorBrewer: ColorBrewer palettes. 2014 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=RColorBrewer.
Nguyen HD, Yoshihama M, Kenmochi N. Phase distribution of spliceosomal introns: implications for intron origin. BMC Evol Biol. 2006:6:69. https://doi.org/10.1186/1471-2148-6-69.
Nguyen L-T, Schmidt HA, Haeseler V, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:32(1):268–274. https://doi.org/10.1093/molbev/msu300.
Nigon V. Les modalites de la reproduction et le determinisme due sexe chez quelques nematodes libres. Ann Sci Nat. 1949:11:1–132.
Noble LM, Yuen J, Stevens L, Moya N, Persaud R, Moscatelli M, Jackson JL, Zhang G, Chitrakar R, Baugh LR, et al. Selfing is the safest sex for *Caenorhabditis tropicalis*. eLife. 2021:10:e62587. https://doi.org/10.7554/eLife.62587.
Nordborg M, Charlesworth B, Charlesworth D. The effect of recombination on background selection. Genet Res. 1996:67(2):159–174. https://doi.org/10.1017/S0016672300033619.
Nordborg M, Donnelly P. The coalescent process with selfing. Genetics. 1997:146(3):1185–1195. https://doi.org/10.1093/genetics/146.3.1185.
Olthof AM, Schwoerer CF, Girardini KN, Weber AL, Doggett K, Mieruszynski S, Heath JK, Moore TE, Biran J, Kanadia RN. Taxonomy of introns and the evolution of minor introns. Nucleic Acids Res. 2024:52(15):9247–9261. https://doi.org/10.1093/nar/gkae550.
Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004:20(2):289–290. https://doi.org/10.1093/bioinformatics/btg412.
Parkinson J, Mitreva M, Whitton C, Thomson M, Daub J, Martin J, Schmid R, Hall N, Barrell B, Waterston RH, et al. A transcriptomic analysis of the phylum nematoda. Nat Genet. 2004:36(12):1259–1267. https://doi.org/10.1038/ng1472.
Pedersen TL. patchwork: the composer of plots. 2024 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=patchwork.
Pettitt J, Philippe L, Sarkar D, Johnston C, Gothe HJ, Massie D, Connolly B, Müller B. Operons are a conserved feature of nematode genomes. Genetics. 2014:197(4):1201–1211. https://doi.org/10.1534/genetics.114.162875.
Pires-daSilva A. Evolution of the control of sexual identity in nematodes. Semin Cell Dev Biol. 2007:18(3):362–370. https://doi.org/10.1016/j.semcdb.2006.11.014.
Pollak E. On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics. 1987:117(2):353–360. https://doi.org/10.1093/genetics/117.2.353.
Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010:26(6):841–842. https://doi.org/10.1093/bioinformatics/btq033.
R Core Team. 2018. R: a language and environment for statistical computing. *Vienna, Austria. www. R-project. org* [Internet]. Available from: https://www.R-project.org/

Reinke V, Cutter AD. Germline expression influences operon organization in the *Caenorhabditis elegans* genome. Genetics. 2009:181(4): 1219–1228. https://doi.org/10.1534/genetics.108.099283.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21(1):245. https://doi.org/10.1186/s13059-020-02134-9.

Robertson HM. Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res. 1998:8(5):449–463. https://doi.org/10.1101/gr.8.5.449.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011:12:77. https://doi.org/10.1186/1471-2105-12-77.

Robinson D, Hayes A, Couch S. broom: convert statistical objects into Tidy Tibbles. 2021 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=broom.

Rockman MV, Kruglyak L. Recombinational landscape and population genomics of *Caenorhabditis elegans*. PLoS Genet. 2009:5(3): e1000419. https://doi.org/10.1371/journal.pgen.1000419.

Roddy AB, Alvarez-Ponce D, Roy SW. Mammals with small populations do not exhibit larger genomes. Mol Biol Evol. 2021:38(9): 3737–3741. https://doi.org/10.1093/molbev/msab142.

Rogozin IB, Carmel L, Csuros M, Koonin EV. Origin and evolution of spliceosomal introns. Biol Direct. 2012:7:11. https://doi.org/10.1186/1745-6150-7-11.

Rosenblad MA, Abramova A, Lind U, Ólason P, Giacomello S, Nystedt B, Blomberg A. Genomic characterization of the barnacle *Balanus* improvisus reveals extreme nucleotide diversity in coding regions. Mar Biotechnol (NY). 2021:23(3):402–416. https://doi.org/10.1007/s10126-021-10033-8.

Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, Baird SE, Haag ES. *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. 2011:7(7):e1002174. https://doi.org/10.1371/journal.pgen.1002174.

Roy PJ, Stuart JM, Lund J, Kim SK. Chromosomal clustering of muscle-expressed genes in *Caenorhabditis elegans*. Nature. 2002:418(6901):975–979. https://doi.org/10.1038/nature01012.

Roy SW, Fedorov A, Gilbert W. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci U S A. 2003:100(12):7158–7162. https://doi.org/10.1073/pnas.1232297100.

Roy SW, Gilbert W. The pattern of intron loss. Proc Natl Acad Sci U S A. 2005:102(3):713–718. https://doi.org/10.1073/pnas.0408274102.

Roy SW, Penny D. Smoke without fire: most reported cases of intron gain in nematodes instead reflect intron losses. Mol Biol Evol. 2006:23(12):2259–2262. https://doi.org/10.1093/molbev/msl098.

Sánchez-Ramírez S, Weiss JG, Thomas CG, Cutter AD. Widespread misregulation of inter-species hybrid transcriptomes due to sex-specific and sex-chromosome regulatory evolution. PLoS Genet. 2021:17(3): e1009409. https://doi.org/10.1371/journal.pgen.1009409.

Sands B, Yun S, Mendenhall AR. Introns control stochastic allele expression bias. Nat Commun. 2021:12(1):6527. https://doi.org/10.1038/s41467-021-26798-4.

Saxena AS, Salomon MP, Matsuba C, Yeh S-D, Baer CF. Evolution of the mutational process under relaxed selection in *Caenorhabditis elegans*. Mol Biol Evol. 2019:36(2):239–251. https://doi.org/10.1093/molbev/msy213.

Shiimori M, Inoue K, Sakamoto H. A specific set of exon junction Complex subunits is required for the nuclear retention of unspliced RNAs in *Caenorhabditis elegans*. Mol Cell Biol. 2013:33(2): 444–456. https://doi.org/10.1128/MCB.01298-12.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005:15(8):1034–1050. https://doi.org/10.1101/gr.3715005.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015: 31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Simpson JT, Durbin R. Efficient de novo assembly of large genomes using compressed data structures. Genome Res. 2012:22(3): 549–556. https://doi.org/10.1101/gr.126953.111.

Sivasundar A, Hey J. Population genetics of *Caenorhabditis elegans*: the paradox of low polymorphism in a widespread species. Genetics. 2003:163(1):147–157. https://doi.org/10.1093/genetics/163.1.147.

Smith MJ, Haigh J. The hitch-hiking effect of a favourable gene. Genet Res (Camb). 1974:23(1):23–35. https://doi.org/10.1017/S0016672300014634.

Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 2003:1(2):e45. https://doi.org/10.1371/journal.pbio.0000045.

Stevens L, Félix M, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T, et al. Comparative genomics of 10 new *Caenorhabditis* species. Evol Lett. 2019:3(2):217–236. https://doi.org/10.1002/evl3.110.

Stevens L, Martínez-Ugalde I, King E, Wagah M, Absolon D, Bancroft R, Gonzalez De La Rosa P, Hall JL, Kieninger M, Kloch A, et al. Ancient diversity in host-parasite interaction genes in a model parasitic nematode. Nat Commun. 2023:14(1):7776. https://doi.org/10.1038/s41467-023-43556-w.

Stevens L, Rooke S, Falzon LC, Machuka EM, Momanyi K, Murungi MK, Njoroge SM, Odinga CO, Ogendo A, Ogola J, et al. The genome of *Caenorhabditis bovis*. Curr Biol. 2020:30(6):1023–-1031.e4. https://doi.org/10.1016/j.cub.2020.01.074.

Sudhaus W, Kiontke K. Comparison of the cryptic nematode species *Caenorhabditis brenneri* sp. n. and *C. remanei* (Nematoda: rhabditidae) with the stem species pattern of the *Caenorhabditis Elegans* group. Zootaxa. 2007:1456(1):45–62. https://doi.org/10.11646/zootaxa.1456.1.2.

Sun S, Kanzaki N, Dayi M, Maeda Y, Yoshida A, Tanaka R, Kikuchi T. The compact genome of *Caenorhabditis niphades* n. sp., isolated from a wood-boring weevil, *Niphades variegatus*. BMC Genomics. 2022:23(1):765. https://doi.org/10.1186/s12864-022-09011-8.

Tandonnet S, Koutsovoulos GD, Adams S, Cloarec D, Parihar M, Blaxter ML, Pires-daSilva A. Chromosome-wide evolution and sex determination in the three-sexed nematode *Auanema rhodensis*. G3 (Bethesda). 2019:9(4):1211–1230. https://doi.org/10.1534/g3.119.0011.

Teterina AA, Willis JH, Lukac M, Jovelin R, Cutter AD, Phillips PC. Genomic diversity landscapes in outcrossing and selfing *Caenorhabditis* nematodes. PLoS Genet. 2023:19(8):e1010879. https://doi.org/10.1371/journal.pgen.1010879.

Teterina AA, Willis JH, Phillips PC. Chromosome-level assembly of the *Caenorhabditis remanei* genome reveals conserved patterns of nematode genome organization. Genetics. 2020:214(4): 769–780. https://doi.org/10.1534/genetics.119.303018.

The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998:282(5396):2012–2018. https://doi.org/10.1126/science.282.5396.2012.
Thomas CG, Li R, Smith HE, Woodruff GC, Oliver B, Haag ES. Simplification and desexualization of gene expression in self-fertile nematodes. Curr Biol. 2012a:22(22):2167–2172. https://doi.org/10.1016/j.cub.2012.09.038.
Thomas CG, Woodruff GC, Haag ES. Causes and consequences of the evolution of reproductive mode in *Caenorhabditis* nematodes. Trends Genet. 2012b:28(5):213–220. https://doi.org/10.1016/j.tig.2012.02.007.
Thomas JH. Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals striking regional cluster domains. Genetics. 2006:172(1):127–143. https://doi.org/10.1534/genetics.104.040030.
Thomas JH, Robertson HM. The *Caenorhabditis* chemoreceptor gene families. BMC Biol. 2008:6:42. https://doi.org/10.1186/1741-7007-6-42.
Tuszynski J. caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc. 2019 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=caTools.
Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Blaxter M, Mieszkowska N, Hall N, Holland P, Durbin R, Richards T, Kersey P, et al. Mitohifi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinformatics. 2023:24(1):288. https://doi.org/10.1186/s12859-023-05385-y.
Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017:27(5):737–746. https://doi.org/10.1101/gr.214270.116.
Venables WN, Ripley BD. Modern applied statistics with S. Fourth. New York: Springer; 2002.
Villeneuve AM. A cis-acting locus that promotes crossing over between X chromosomes in *Caenorhabditis elegans*. Genetics. 1994:136(3):887–902. https://doi.org/10.1093/genetics/136.3.887.
Vinogradov AE. Testing genome complexity. Science. 2004:304(5669):389–390. https://doi.org/10.1126/science.304.5669.389b.
Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017:33(14):2202–2204. https://doi.org/10.1093/bioinformatics/btx153.
Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009:136(5669):701–718. https://doi.org/10.1016/j.cell.2009.02.009.
Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014:9:e112963. https://doi.org/10.1371/journal.pone.0112963.
Wang JR, Holt J, McMillan L, Jones CD. FMLRC: hybrid long read error correction using an FM-index. BMC Bioinformatics. 2018:19(1):50. https://doi.org/10.1186/s12859-018-2051-3.
Wei T, Simko V. R package "corrplot": visualization of a correlation matrix. 2021 [accessed 2024 Jul 1]. https://github.com/taiyun/corrplot.
Whitney KD, Boussau B, Baack EJ, Garland T. Drift and genome complexity revisited. PLoS Genet. 2011:7(6):e1002092. https://doi.org/10.1371/journal.pgen.1002092.
Whitney KD, Garland T. Did genetic drift drive increases in genome complexity? PLoS Genet. 2010:6(8):e1001080. https://doi.org/10.1371/journal.pgen.1001080.
Wickham H. Reshaping data with the reshape package. J Stat Softw. 2007:21(12):1–20. https://doi.org/10.18637/jss.v021.i12.
Wickham H. tidyr: tidy messy data. 2021 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=tidyr.
Wickham H, François R, Henry L, Müller K, RStudio. dplyr: a grammar of data manipulation. 2022 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=dplyr.
Wickham H, Pedersen TL. gtable: arrange "grobs" in tables. 2019 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=gtable.
Wickham H, Sievert C. Ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009.
Woodruff GC, Eke O, Baird SE, Félix M-A, Haag ES. Insights into species divergence and the evolution of hermaphroditism from fertile interspecies hybrids of *Caenorhabditis* nematodes. Genetics. 2010:186(3):997–1012. https://doi.org/10.1534/genetics.110.120550.
Woodruff GC, Teterina AA. Degradation of the repetitive genomic landscape in a close relative of *Caenorhabditis elegans*. Mol Biol Evol. 2020:37(9):2549–2567. https://doi.org/10.1093/molbev/msaa107.
Woollard A. Gene duplications and genetic redundancy in *C. elegans*. *WormBook*. 2005 [accessed 2024 Jul 1]. http://www.wormbook.org/chapters/www_geneduplication/geneduplication.html.
Xie D, Ye P, Ma Y, Li Y, Liu X, Sarkies P, Zhao Z. Genetic exchange with an outcrossing sister species causes severe genome-wide dysregulation in a selfing *Caenorhabditis* nematode. Genome Res. 2022:32(11-12):2015–2027. https://doi.org/10.1101/gr.277205.122.
Xie Y, Ren Y. Mechanisms of nuclear mRNA export: a structural perspective. Traffic. 2019:20(11):829–840. https://doi.org/10.1111/tra.12691.
Xu G-C, Xu T-J, Zhu R, Zhang Y, Li S-Q, Wang H-W, Li J-T. LR_gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience. 2019:8(1):giy157. https://doi.org/10.1093/gigascience/giy157.
Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L, et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience. 2020:9(9):giaa094. https://doi.org/10.1093/gigascience/giaa094.
Ye C, Ji G, Liang C. detectMITE: a novel approach to detect miniature inverted repeat transposable elements in genomes. Sci Rep. 2016:6:19688. https://doi.org/10.1038/srep19688.
Yin D, Schwarz EM, Thomas CG, Felde RL, Korf IF, Cutter AD, Schartner CM, Ralston EJ, Meyer BJ, Haag ES. Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science. 2018:359(6371):55–61. https://doi.org/10.1126/science.aao0827.
Yu G. ggplotify: Convert plot to "grob" or "ggplot" object. 2021 [accessed 2024 Jul 1]. https://CRAN.R-project.org/package=ggplotify.
Yu G, Smith DK, Zhu H, Guan Y, Lam TT. Ggtree : an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017:8(1):28–36. https://doi.org/10.1111/2041-210X.12628.
Zhang G, Wang Y, Andersen EC. Natural variation in *C. elegans* short tandem repeats. Genome Res. 2022:32(10):1852–1861. https://doi.org/10.1101/gr.277067.122.
Zheleva A, Gómez-Orte E, Sáenz-Narciso B, Ezcurra B, Kassahun H, De Toro M, Miranda-Vizuete A, Schnabel R, Nilsen H, Cabello J. Reduction of mRNA export unmasks different tissue sensitivities to low mRNA levels during *Caenorhabditis elegans* development. PLoS Genet. 2019:15(9):e1008338. https://doi.org/10.1371/journal.pgen.1008338.

Associate editor: Esther Betran