Chromosome-Level Assembly of the *Caenorhabditis remanei* Genome Reveals Conserved Patterns of Nematode Genome Organization

Anastasia A. Teterina,*,† John H. Willis,* and Patrick C. Phillips*,1

*Institute of Ecology and Evolution, University of Oregon, Eugene, Oregon 97403 and †Center of Parasitology, A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow 117071, Russia

ORCID IDs: 0000-0003-3016-8720 (A.A.T.); 0000-0001-7271-342X (P.C.P.)

ABSTRACT The nematode *Caenorhabditis elegans* is one of the key model systems in biology, including possessing the first fully assembled animal genome. Whereas *C. elegans* is a self-reproducing hermaphrodite with fairly limited within-population variation, its relative *C. remanei* is an outcrossing species with much more extensive genetic variation, making it an ideal parallel model system for evolutionary genetic investigations. Here, we greatly improve on previous assemblies by generating a chromosome-level assembly of the entire *C. remanei* genome (124.8 Mb of total size) using long-read sequencing and chromatin conformation capture data. Like other fully assembled genomes in the genus, we find that the *C. remanei* genome displays a high degree of synteny with *C. elegans* despite multiple within-chromosome rearrangements. Both genomes have high gene density in central regions of chromosomes relative to chromosome ends and the opposite pattern for the accumulation of repetitive elements. *C. elegans* and *C. remanei* also show similar patterns of interchromosome interactions, with the central regions of chromosomes appearing to interact with one another more than the distal ends. The new *C. remanei* genome presented here greatly augments the use of the *Caenorhabditis* as a platform for comparative genomics and serves as a basis for molecular population genetics within this highly diverse species.

KEYWORDS *Caenorhabditis remanei*; *Caenorhabditis elegans*; chromosome-level assembly; comparative genomics

THE free-living nematode *Caenorhabditis elegans* is one of the most-used and best-studied model organisms in genetics, developmental biology, and neurobiology (Brenner 1973, 1974; Blaxter 1998). *C. elegans* was the first multicellular organism with a complete genome sequence (*C. elegans* Sequencing Consortium 1998), and the *C. elegans* genome currently has one of the best-described functional annotations among metazoans, as well as possessing hundreds of large-scale data sets focused on functional genomics (Gerstein *et al.* 2010). The genome of *C. elegans* is compact,

roughly 100 Mb [100.4 Mb is the "classic" N2 assembly (*C. elegans* Sequencing Consortium 1998); 102 Mb is the V2010 strain genome (Yoshimura *et al.* 2019)], and consists of six holocentric chromosomes, five of which are autosomes and one that is a sex chromosome (*X*). All chromosomes of *C. elegans* have a similar pattern of organization: a central region occupying about one-half of the chromosome that has a low recombination rate, low transposon density, and high gene density, and the "arms" display the characteristics exactly opposite to this (Waterston *et al.* 1992; Barnes *et al.* 1995; Rockman and Kruglyak 2009).

About 65 species of the *Caenorhabditis* genus are currently known (Kiontke *et al.* 2011), and for many of them genomic sequences are available (Stevens *et al.* 2019) (http://www.wormbase.org/ and https://evolution.wormbase.org). Most of the *Caenorhabditis* nematodes are outcrossing species with females and males (gonochoristic), but three species — *C. elegans*, *C. briggsae*, and *C. tropicalis* — reproduce primarily

via self-fertilizing ("selfing") hermaphrodites with rare males (androdioecy) (Kiontke *et al.* 2011). *Caenorhabditis* species have the *XX/X*O sex determination: females and hermaphrodites carry two copies of the *X* chromosomes, while males have only one *X* chromosome (Pires-daSilva 2007).

C. remanei is an obligate outcrossing nematode, a member of the "Elegans" supergroup, which has become an important model for natural variation (Jovelin *et al.* 2003; Reynolds and Phillips 2013), experimental evolution (Sikkink *et al.* 2014a, 2015, 2019; Castillo *et al.* 2015), and population genetics (Graustein *et al.* 2002; Cutter and Charlesworth 2006; Cutter *et al.* 2006; Dolgin *et al.* 2007; Jovelin *et al.* 2009; Dey *et al.* 2012). Whole-genome data are available for three strains of *C. remanei* (Table 1), but all of these assemblies are fragmented. To improve genomic precision for experimental studies and to facilitate the analysis of chromosome-wide patterns of genome organization, recombination, and diversity, the complete assembly for this species is required. We generated a chromosome-level genome assembly of the *C. remanei* PX506 inbred strain using a long-read/Hi-C approach, and used this new chromosome-level resolution in a comparative framework to reveal global similarities in genome organization and spatial chromosome interactions between *C. elegans* and *C. remanei*.

Materials and Methods

Nematode strains

Nematodes were maintained under standard laboratory conditions as described previously (Brenner 1974). *C. remanei* isolates were originally derived from individuals living in association with terrestrial isopods (family *Oniscidea*) collected from Koffler Scientific Reserve at Jokers Hill, King City, Toronto, Ontario, as described in Sikkink *et al.* (2014b). Strain PX393 was founded from a cross between single female and male *C. remanei* individuals isolated from isopod Q12. This strain was propagated for two to three generations before freezing. PX506, the source of the genome described here, is an inbred strain derived from PX393 following sibling mating for 30 generations to reduce within-strain heterozygosity. This strain was frozen and subsequently recovered at large population size for several generations before further experimental analysis.

Sequencing and genome assembly of the C. remanei reference

Strain PX506 was grown on 20 110-mm plates until its entire *Escherichia coli* food source (strain OP50) was consumed. Worms were washed 5× in M9 using 15-ml conical tubes and spun at a low speed to concentrate. The worm pellet was flash frozen and genomic DNA was isolated using a Genomic-tip 100/G column (QIAGEN, Valencia, CA). Next, 4 µg (average size 23 kb) was frozen and shipped to Dovetail Genomics (Santa Cruz, CA; https://dovetailgenomics.com), along with frozen whole animals for subsequent Pacific Biosciences (PacBio) and Hi-C analysis. The *C. remanei* PX506

inbred strain was sequenced and assembled by Dovetail Genomics. The primary contigs were generated from two PacBio single-molecule real-time (SMRT) Cells using the FALCON assembly (Chin *et al.* 2016) followed by Arrow polishing (https://github.com/PacificBiosciences/GenomicConsensus). The final scaffolds were constructed with Dovetail Genomics Hi-C library sequences and the HiRise software pipeline (Putnam *et al.* 2016). Additionally, we performed whole-genome sequencing of the PX506 strain using the Nextera kit (Illumina) for 100-bp paired-end read sequencing on the Illumina Hi-Seq 4000 platform (University of Oregon Sequencing Facility, Eugene, OR).

We then performed a BLAST (Basic Local Alignment Search Tool) search (Altschul *et al.* 1990) against the National Center for Biotechnology Information (NCBI) GenBank nucleotide database (Benson *et al.* 2012) and filtered any scaffolds (E-value <1e−15) of bacterial origin. Short scaffolds with good matches to *Caenorhabditis* nematodes were aligned to six chromosome-sized scaffolds by GMAP v.2018-03-25 (Wu and Watanabe 2005) and visualized in IGV v.2.4.10 (Thorvaldsdóttir *et al.* 2013) to examine whether they represent alternative haplotypes.

The final filtered assembly was compared to the "recompiled" version of the *C. elegans* reference genome generated from strain VC2010, a modern strain derived from the classical N2 strain (Yoshimura *et al.* 2019), and *C. briggsae* genomes (available under accession numbers PRJEB28388 from the NCBI Genome database and PRJNA10731 from WormBase WS260) by MUMmer3.0 (Kurtz *et al.* 2004). The names and orientations of the *C. remanei* chromosomes were defined by the longest total nucleotide matches in proper orientation to *C. elegans* chromosomes. Dot plots with these alignments were plotted using the ggplot2 package (Wickham 2016) in R (R Core Team 2018). The completeness of the *C. remanei* genome assembly was assessed by BUSCO v.3.0.2 (Simão *et al.* 2015) with the Metazoa odb9 and Nematoda odb9 databases. Results were visualized with generate_plot_xd_v2.py script (https://github.com/xieduo7/my_script/blob/master/busco_plot/generate_plot_xd_v2.py).

The mitochondrial genome was generated using a reference mitochondrial genome of *C. remanei* (KR709159.1) from the NCBI database (http://www.ncbi.nlm.nih.gov/nucleotide/) and Illumina reads of the *C. remanei* PX506 inbred strain. The reads were aligned with bwa mem v.0.7.17 (Li and Durbin 2009), filtered with samtools v.1.5 (Li *et al.* 2009a). We marked PCR duplicates in the mitochondrial assembly with MarkDuplicates from picard-tools v.2.0.1 (http://broadinstitute.github.io/picard/), realigned insertions/deletions and called variants with IndelRealignment and HaplotypeCaller in the haploid mode from GATK tools v.3.7 (McKenna *et al.* 2010), filtered low-quality sites, and then used bcftools consensus v.1.5 (Li 2011) to generate the new reference mitochondrial genome. To estimate the residual heterozygosity throughout the rest of the genome, we implemented a similar read-mapping protocol but used the default parameters to call genotypes and then filtered

variants using standard hard filters (residual_heterozygosity.sh and plot_residual_heterozygosity.R).

Repeat masking in C. remanei and C. elegans

For repeat masking, we created a comprehensive repeat library (Coghlan et al. 2018; see also instructions at http://avrilomics.blogspot.com) and masked sequence-specific repeat motifs, as described in Woodruff and Teterina (2019 preprint). De novo repeat discovery was performed by RepeatModeler v.1.0.11 (Smit and Hubley 2008) with the NCBI engine. Transposon elements were detected by transposonPSI (http://transposonpsi.sourceforge.net), with sequences shorter than 50 bases filtered out. Inverted transposon elements were located with detectMITE v.2017-04-25 (Ye et al. 2016) with default parameters. Transfer RNAs were identified with tRNAscan-SE v.1.3.1 (Lowe and Eddy 1997) and their sequences were extracted from a reference genome by the getfasta tool from the BEDTools package v.2.25.0 (Quinlan and Hall 2010). We searched for LTR retrotransposons as described at http://avrilomics.blogspot.com/2015/09/ltrharvest.html, by LTRharvest and LTRdigest from GenomeTools v.1.5.11 (Gremme et al. 2013) with domains from the Gypsy Database (Llorens et al. 2010), and several models of Pfam protein domains (Finn et al. 2015), listed in Tables SB1 and SB2 of Steinbiss et al. (2009). To filter LTRs, we used two scripts: https://github.com/satta/ltrsift/blob/master/filters/filter_protein_match.lua and https://gist.github.com/avrilcoghlan/4037d6b8cca32eaf48b0.

Additionally, we uploaded nematode repeats from the Dfam database (Hubley et al. 2015) using the queryRepeatDatabase.pl script from the RepeatMasker v.4.0.7 (Smit et al. 2015) utilities with the "–species rhabditida" option, and C. elegans and ancestral repetitive sequences from Repbase v.23.03, (Bao et al. 2015). We then combined all repetitive sequences obtained from these tools and databases in one redundant repeat library. We clustered those sequences with < 80% identity by uclust from the USEARCH package v.8.0, (Edgar 2010) and classified them via the RepeatMasker Classify tool v.4.0.7, (Smit et al. 2015). Potential protein matches with C. remanei (PRJNA248911) or C. elegans protein sequences (PRJNA13758) from WormBase W260 were detected with BLASTX (Altschul et al. 1990). The repetitive sequences classified as "unknown" and having BLAST hits with E-value ≤ 0.001 with known protein-coding genes were removed from the final repeat libraries.

For C. remanei, the final repeat library was used by RepeatMasker with "–s" and "–gff" options. An additional round of masking was performed with the "–species caenorhabditis" option. The genome was also masked with the redundant repeat library acquired before the clustering step. Regions that were masked with the redundant library but not masked with the final library were extracted using BEDTools subtract and classified by RepeatMasker Classify. Additionally, we checked the depth coverage with the Illumina reads in these regions, as regions classified as a known type of repeat and displaying coverage > 70 were masked in the reference

genome by BEDTools maskfasta. The masked regions were extracted to a bed file with a bash script (https://gist.github.com/danielecook/cfaa5c359d99bcad3200), and the same regions were soft masked by BEDTools maskfasta with the "–soft" option.

Using the same approach, we masked the C. elegans reference N2 strain (PRJNA13758 from WormBase W260) and then extracted all regions that were masked in the "official" masked version of the genome but not masked by our final repeat library. These regions were extracted, classified by RepeatMasker with default parameters, and searched against C. elegans proteins with the BLASTX algorithm and the C. elegans reference genome with BLASTN. Regions with unknown class and a match with C. elegans proteins (see above) were removed. Regions with > 5 matches and an E-value ≤ 0.001 with the C. elegans genome were added to the final database, and used to mask the C. elegans reference genome generated from strain VC2010 (Yoshimura et al. 2019). The same regions were soft-masked with BEDTools maskfasta.

Full-length transcript sequencing

We used single-molecule long-read RNA sequencing (Iso-Seq) to obtain high-quality transcriptomic data. We used the Clonetech SMARTer PCR complementary DNA (cDNA) Synthesis kit for cDNA synthesis and PCR amplification with no size selection starting with 500 ng of total RNA from a mixed-staged population of C. remanei strain PX506 (Cat#634925; Clonetech). PacBio library generation was performed on-site at the University of Oregon Genomics and Cell Characterization Core Facility and sequenced on a PacBio Sequel I platform utilizing four SMRT cells of data.

We generated circular consensus reads using the ccs tool with "–noPolish –minPasses 1" options from PacBio SMRT link tools v.5.1.0 (https://www.pacb.com/support/software-downloads/) and obtained full-length transcripts with lima from the same package with "–isoseq –no-pbi" options. Next, trimmed reads from all SMRT cells were merged together, clustered, and polished with isoseq3 tools v.3.2 (https://github.com/PacificBiosciences/IsoSeq), and mapped to the C. remanei reference genome with GMAP. Redundant isoforms were collapsed by collapse_isoforms_by_sam.py from Cupcake ToFU (https://github.com/Magdoll/cDNA_Cupcake). The longest ORFs were predicted with TransDecoder v.5.0.1 (Haas et al. 2013) and used as coding sequence (CDS) hints in the genome annotation (see below).

Genome annotation

We performed de novo annotation of the C. remanei genome using the following hybrid approach. For ab initio gene prediction, we applied the GeneMark-ES algorithm v.4.33 (Ter-Hovhannisyan et al. 2008) with default parameters. De novo gene prediction with the MAKER pipeline v.2.31.9 (Holt and Yandell 2011) was carried out with C. elegans (PRJNA13758), C. briggsae (PRJNA10731), and C. latens (PRJNA248912) proteins from WormBase 260, excluding the repetitive regions

identified above. To implement gene prediction using the BRAKER pipeline v.2.1.0 (https://github.com/Gaius-Augustus/BRAKER), we included RNA-sequencing (RNA-seq) from our previous *C. remanei* studies (SRX3014311 and SRP049403).

Annotations from BRAKER2, MAKER2, and GeneMark-ES were combined in EVidenceModeler v.1.1.1 (Haas *et al.* 2008) with weights 6, 3, and 1, correspondingly. CDS from the EVidenceModeler results were used to train AUGUSTUS version 3.3 (Stanke *et al.* 2006) as described on http://bioinf.uni-greifswald.de/augustus/binaries/tutorial/training.html. Next, models were optimized and retrained again, then we created a file with extrinsic information with factor 1000 and malus 0.7 for CDS, and all other options as in "extrinsic.E.cfg" for annotation with est database hits from the AUGUSTUS supplemental files. The final annotation was executed with Iso-Seq data as the hints file and EVidenceModeler -trained models with "–singlestrand=true –gff3=on –UTR=off". Scanning for known protein domains and the functional annotation were conducted with InterProScan v.5.27-66.0 (Quevillon *et al.* 2005). We validated/filtered final gene models according to coverage with RNA-seq and Iso-Seq data, matches with known *Caenorhabditis* proteins, and protein/transposon domains.

We identified one-to-one orthologs of *C. remanei* and *C. elegans* proteins using orthofinder2 (Emms and Kelly 2019); for *C. elegans* we used only proteins validated in the VC2010 (Yoshimura *et al.* 2019). The identities of the proteins were estimated by pairwise global alignments using calc_pc_id_between_seqs.pl script (https://gist.github.com/avrilcoghlan/5311008). Gene synteny plots were made in R with a custom script (synteny_plot.R).

Genome activity and features

We studied patterns of genome activity in *C. elegans* and *C. remanei* using *C. elegans* Hi-C data from the (Brejc *et al.* 2017) study (SRR5341677–SRR5341679) and the Hi-C reads produced in the current study, as well as available RNA-seq data from the L1 larval stage for *C. elegans* (SRR016680, SRR016681, and SRR016683) and *C. remanei* (SRP049403). Hi-C reads were mapped to the reference genomes with bwa mem and RNA-seq read with STAR v.2.5 (Dobin *et al.* 2013) using the default parameters and gene annotations; to count reads for transcripts, we used htseq-count from the HTSeq package v.0.9.1 (Anders *et al.* 2015) and corrected by the total lengths of the gene CDSs [reads per kilobase of transcript per million mapped reads (RPKM)] using a bash script (https://gist.github.com/darencard/fcb32168c243b92734e85c5f8b59a1c3) and a custom R script (RNA_seq_R_analysis_and_figures.R). For Hi-C interactions, we applied the Arima pipeline (https://github.com/ArimaGenomics/mapping_pipeline), BEDTools, and a custom bash script (Hi-C_analysis_with_ARIMA.sh and Hi-C_R_analysis_and_figures.R).

We calculated the fraction of exonic/intronic DNA and the number of genes per 100-kb windows from the genome annotations using the BEDtools coverage tool. GC content and the percent of repetitive regions were estimated, correspondingly, from the unmasked and hard-masked genomes via BEDtools nuc, also on 100-kb windows by a custom script (get_genomic_fractions.sh). For the formal statistical tests, we defined chromosome "centers" to be the central one-half of a chromosome and the "arms" to be the peripheral one-quarter of each length on either side of the center. To measure the positional effect of these genomic features, we conducted the Cohen's d effect size test with package "lsr" in R (Navarro 2013) and calculated statistical differences using the Wilcoxon–Mann–Whitney test using basic R (see a custom script fractions_stats_and_figures.R).

Data availability

Strain PX506 is available from the *Caenorhabditis* Genetic Center. All raw sequencing data generated in this study have been submitted to the NCBI BioProject database (https://www.ncbi.nlm.nih.gov/bioproject/) under accession number PRJNA577507. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession WUAV00000000; the version described in this paper is version WUAV01000000. The reference genome assembly is available at the NCBI Genome database (https://www.ncbi.nlm.nih.gov/genome/) under accession number GCA_010183535.1. Supplementary custom scripts to estimate statistics, and generate main and supplemental figures, are available on GitHub (https://github.com/phillips-lab/C.remanei_genome). Supplemental material available at figshare: https://doi.org/10.25386/genetics.11889099. All online resources mentioned in the manuscript were accessed in March 2020.

Results

New reference genome assembly and annotation

We generated a high-quality chromosome-level assembly of the *C. remanei* PX506 inbred line with deep PacBio whole-genome sequencing (∼100× coverage by 1.3 million reads) and Hi-C (∼900× with 418 million paired-end Illumina reads). Assembly of the PacBio sequences resulted in 135.85 Mb of genome and bacterial sequences with 298 scaffolds. The Hi-C data dramatically improved the PacBio assembly, and the HiRise scaffolding increased the N50 from 4.042 to 21.502 Mb by connecting scaffolds from the PacBio assembly together, resulting in 235 scaffolds (see the summary statistics in Table 1 and Supplemental Material, Tables S1–3). After the filtering of scaffolds of bacterial origin, six chromosome-sized scaffolds were obtained, as expected (Figure S1). Additionally, there were 180 short scaffolds that are alternative haplotypes or unplaced scaffolds (the average length is 31,169 nt with SD of 48,700 and a median length of 19,076 nt). Because only the long-sized fraction of total DNA was selected in the long-read library, the mitochondrial DNA was not covered by PacBio sequencing. The mitochondrial genome was therefore generated independently using the Illumina whole-genome data of the reference strain (see

Table 1 Available genome assemblies of C. remanei

Strain	NCBI ID	Total size (Mb)	Number of scaffolds	Scaffold N50 (Mb)	Scaffold L50	GC%	Number of genes
PB4611	GCA_000149515.1	145.443	3,670	0.435	70	38.50	32,412
PX356	GCA_001643735.2	118.549	1,591	1.522	10	35.90	24,977
PX439	GCA_002259225.1	124.542	912	1.765	13	35.30	24,867
PX506[a]	GCA_010183535.1	124.870	7	21.502	3	37.96	26,308

NCBI ID, National Center for Biotechnology Information identifier.
[a] This study.

Materials and Methods). The total length of the new *C. remanei* reference genome without alternative haplotypes is 124,870,449 bp, which is very close in size to previous assemblies of other *C. remanei* strains (Table 1). After 30 generations of inbreeding, the residual heterozygosity of the PX506 line remained at 0.02% of SNPs (a 100-fold decrease relative to population-level variability; Dey *et al.* 2012). Most of the remaining polymorphic sites in PX506 are located in the peripheral parts of chromosomes, with one-half of all sites on the *X* chromosome (Figure S2).

To assess the quality of the new reference, we performed a standard BUSCO analysis (Simão *et al.* 2015). The new assembly of PX506 presented here has 975 of 982 BUSCO genes for completeness (97.9% based on the Nematode database) and displays fewer missed and duplicated genes than the previous assembly (PX356), but for the most part the BUSCO scores are very similar (see Figure S3).

We used full-length transcripts, RNA-seq data from previous *C. remanei* studies, known *Caenorhabditis* proteins, and *ab initio* predictions to annotate the *C. remanei* genome (see *Materials and Methods*). The final annotation contains 26,308 protein-coding genes, which is close to the number of annotated genes in other *C. remanei* strains (Table 1). Each of the genes predicted by AUGUSTUS has been validated by at least one type of evidence: 25,380 genes have hits with known *Caenorhabditis* proteins (23,840 with *C. elegans*, *C. briggsae*, and *C. latens*), including 25,373 that have matches with the previously annotated genes of *C. remanei*; 19,285 contain known protein domains or functional annotation; 18,662 were supported by RNA-seq data; and 8870 have full-transcript evidence derived from 19,410 high-quality isoforms from the Iso-Seq data. In addition, 27 genes were predicted from the full-transcript data.

Synteny of C. remanei and C. elegans

We identified 11,160 one-to-one orthologs of *C. remanei* and *C. elegans* protein-coding genes, which, after additional filtering on the global-alignment identity, resulted in 9247 ortholog pairs. Comparison of our new chromosome-level assembly to that of *C. elegans* revealed that the *C. remanei* and *C. elegans* genomes are in high synteny, despite having a very large number of within-chromosome rearrangements (only 120 of ortholog pairs are not located on homologous chromosomes). The distribution of orthologs across chromosomes is fairly uniform (chromosome I contains 1511 orthologs; II contains 1498; III contains 1473; IV contains 1472; V contains 1703; and *X* contains 1470). The central domains of autosomes and most regions of the *X* chromosomes are more highly conserved than the rest of the genome (Figure 1). Orthologs located on the *X* chromosome have greater global identity than ones located on autosomes ($W = 532{,}160$, P-value $= 0.0128$).

We chose the orientation of the *C. remanei* chromosomes based on the same/inverted directions of nucleotide alignments and one-to-one orthologs in *C. elegans*. However, it appears that the ancestral orientation of chromosome III is actually inverted relative to the *C. elegans* standard based on syntenic blocks between *C. briggsae* and *C. remanei* (*e.g.*, *C. elegans* chromosome III has undergone large-scale inversion since divergence from the common ancestor of these three species, see the dot plots in Figure S4).

Organization of C. remanei and C. elegans chromosomes

To compare the genomic organization of *C. remanei* and *C. elegans*, we identified repetitive sequences in the *C. remanei* PX506 genome and the updated reference of *C. elegans* (Yoshimura *et al.* 2019). In total, 22.04% from 124.8 Mb of the *C. remanei* genome and 20.77% from 102 Mb of the *C. elegans* genome were repetitive. All homologous chromosomes of *C. remanei* are, on average, 22% longer than corresponding homologous chromosomes in *C. elegans*; the physical sizes of chromosomes I, II, III, IV, V, and *X* are 15.3, 15.5, 14.1, 17.7, 21.2, and 18.1 Mb in *C. elegans* and 17.2, 19.9, 17.8, 25.7, 22.5, and 21.5 Mb in *C. remanei*, respectively. These findings are consistent with the conclusions of Fierst *et al.* (2015) that the differences in the genome sizes of outcrossing and selfing *Caenorhabditis* species cannot be explained solely by an increase in transposable element abundance.

To identify finer-scale patterns displayed across each chromosome, we estimated fractions of exons, introns, and repetitive DNA per 100-kb windows (Figure 2), as well as GC content, gene counts, and gene fractions (Figure S5). In general, *C. elegans* and *C. remanei* display analogous patterns of organization across all chromosomes. Repetitive DNA was found in greater quantities in the peripheral parts of chromosomes of *C. elegans* (Cohen's d $= 1.58$, $W = 232{,}780$, P-value $< 2.2e{-}16$) and *C. remanei* (Cohen's d $= 1.44$, $W = 332{,}810$, P-value $< 2.2e{-}16$). Repetitive regions of *C. elegans* (VC2010) and *C. remanei* (PX506) genomes are available in Files S3 and S4.

Further, the fractions of the repetitive DNA in both species are negatively correlated with number of genes ($r = -0.26$ in
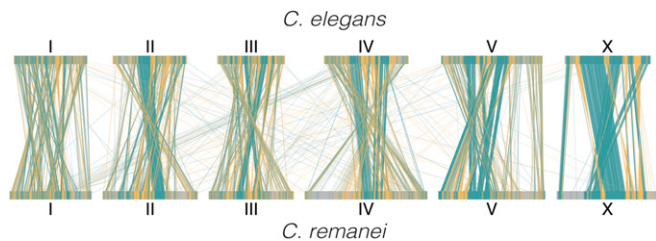
Figure 1 Gene synteny plot of the 9247 one-to-one orthologs of *C. elegans* (the top row) and *C. remanei* (the bottom row). The lines connect locations of the orthologs on the *C. elegans* and *C. remanei* reference genomes. Teal lines represent genes in the same orientation, whereas orange lines show genes in an inverted orientation.

C. elegans and $r = -0.45$ in *C. remanei*) and the exonic fractions ($r = -0.43$ and -0.63). There is an inverse positional effect with respect to the number of genes: more genes are located in the central domain than in the peripheral parts of chromosomes in *C. elegans* (Cohen's d = 0.44, $W = 93,950$, *P*-value = 2.9e−15) and in the *C. remanei* genome (Cohen's d = 0.72, $W = 116,470$, *P*-value < 2.2e−16), as has long been noted in *C. elegans* (Barnes *et al.* 1995). Both species have a similarly high density of genes (211.2 and 216.2 genes per megabase for *C. elegans* and *C. remanei*, respectively), which is one order of magnitude higher than for humans (Dunham *et al.* 2004). Not surprisingly, then, genes occupy a large fraction of the genome in both *C. elegans* (the mean fraction per 100 kb is 0.58, 95% C.I. 0.569–0.5836) and *C. remanei* (the mean fraction equals 0.44, 95% C.I. 0.436–0.452). Genes on the arms have longer total intron sizes then in the central domains (*C. elegans*: $W = 53,932,000$, *P*-value < 2.2e−16; *C. remanei*: $W = 82,764,000$, *P*-value < 2.2e−16). GC content, number genes, and gene fraction also differ between central and peripheral parts of chromosomes, as shown in Figure S5 and Table S4.

In both species, there is more intronic DNA in the peripheries of chromosomes than in their centers (Cohen's d = 0.68, $W = 177,730$, *P*-value < 2.2e−16 for *C. elegans*; Cohen's d = 0.32, $W = 227,220$, *P*-value = 1e−06 for *C. remanei*), although for *C. remanei* this effect is strongly driven by the different distributions of introns on chromosomes IV and X (Figure 2). Overall, 28.5 and 27.3% of total intron lengths consist of repetitive elements in *C. elegans* and *C. remanei*. Additionally, we investigated the transcriptional landscapes of the *C. elegans* and *C. remanei* genomes at the L1 larval stage, and found that the expression of genes in the central domain is very slightly, yet significantly, larger than gene expression in the peripheral domains (Cohen's d = 0.06, $W = 9,796,800$, *P*-value < 2.2e−16 for *C. elegans*; Cohen's d = 0.04, $W = 29,629,000$, *P*-value = 2.9e−14 for *C. remanei*); the chromosome-wise distribution of RPKM is shown in Figure S6.

Similar patterns of within-genome interactions

In examining the pattern of read mapping of Hi-C data across the *C. remanei* genome, we noted that the central domain of each chromosome appears to be enriched for interactions

with the central domains of all other chromosomes (Figure S1). To explore this further, we examined the distances of three-dimensional (3D) interactions within chromosomes and proportions of interchromosomal contacts in *C. remanei* and *C. elegans* genomes. This analysis should be considered preliminary, as the data are likely noisy since they were obtained from mixed tissues and the *C. remanei* sample was collected from mixed developmental stages (including adult worms), whereas the *C. elegans* results are derived from a reanalysis of data from embryos (Brejc *et al.* 2017). At the moment, Hi-C data for different developmental stages of *C. elegans* and/or *C. remanei* are not publicly unavailable.

A total of 12% of the 199.2 million read pairs mapped to different chromosomes of *C. remanei*, which indicates a high level of potential *trans*-chromosome interactions. We observed an even higher proportion (32.7% from 123.9 million read pairs) of *trans*-chromosome contacts in the *C. elegans* sample. When we consider interactions within rather than between chromosomes, we find that the central domains tend to have a larger median distance between interaction pairs compared to the arms. This difference is significant within both species (Figure 3A; Cohen's d = 1.46, $W = 39,418$, *P*-value < 2.2e−16 for *C. elegans*; Cohen's d = 1.74, $W = 45,396$, *P*-value < 2.2e−16 for *C. remanei*).

Central domains tend to interact with other central domains in *C. remanei* (Figure 3A; 36.2% center–center contacts, 40.6% arm–center, and 19.3% arm–arm), but the proportion of center–center contacts in *C. elegans* is lower (27.8% center–center, 49.7% arm–center, and 22.5% arm–arm). The deviation from the expected uniform distribution (one center–center: two center–arm/arm–center: one arm–arm) of *trans*-chromosome interactions is larger in *C. remanei* than in *C. elegans* ($\chi^2 = 330,220$, d.f. = 2, *P*-value < 2.2e−16 for *C. elegans*; $\chi^2 = 1,643,800$, d.f. = 2, *P*-value < 2.2e−16 for *C. remanei*). All chromosomes, both in the *C. elegans* and *C. remanei* samples, have almost even numbers of contacts with other chromosomes (*C. elegans*: chromosome I has 15.1% from all interchromosomal contacts, II has 15.3%, III has 14.2%, IV has 17%, V has 19.5%, and X has 18.8%; *C. remanei*: I has 16.1%, II has 16.7%, III has 16.1%, IV has 18.2%, V has 18.2%, and X has 14.7%). However, if we focus specifically on windows with localized contacts we see that within *C. remanei*, interactions are more dispersed on X and V chromosomes and there are areas of thick contacts in the central parts of autosomes, whereas in the *C. elegans* sample all chromosomes actively interact (Figure 3B).

Discussion

We have generated a high-quality reference genome of the *C. remanei* line PX506, which is now one of the five currently available chromosome-level assemblies of *Caenorhabditis* nematodes of the Elegans supergroup, including two selfing species, *C. elegans* (*C. elegans* Sequencing Consortium 1998) and *C. briggsae* (Stein *et al.* 2003), and outcrossing *C. inopinata* (Kanzaki *et al.* 2018) and *C. nigoni* (Yin *et al.* 2018).
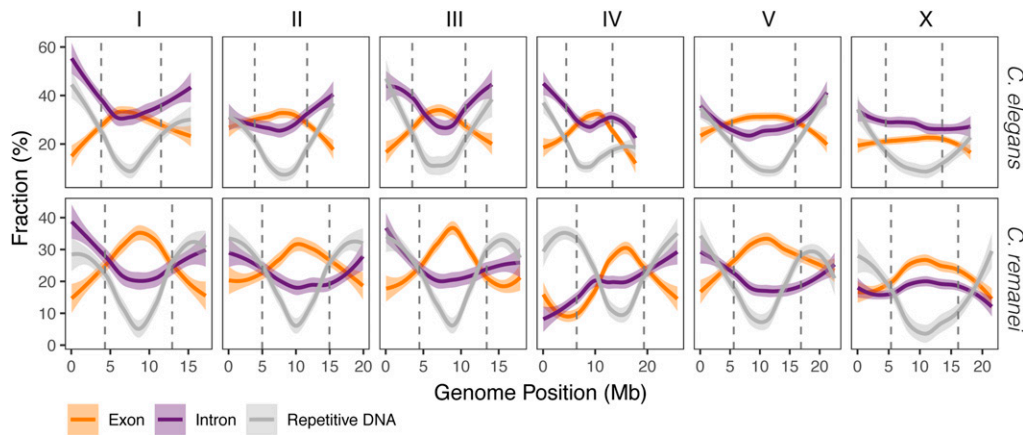
Figure 2 Genomic landscape of genetic elements for *C. elegans* and *C. remanei*. Vertical dashed lines show the boundaries of the central domain. Colored lines represent the smoothed means of the fraction of repetitive DNA (gray), exons (orange), and introns (purple) calculated from 100-kb windows. Shaded areas show 95% C.I.s of the mean.

C. remanei is an outcrossing nematode with high genetic diversity in comparison with *C. elegans*, *C. briggsae*, and *C. tropicalis* (Jovelin *et al.* 2003; Cutter *et al.* 2006). Therefore, to reduce the diversity and improve the quality of assembly, we constructed a highly inbred line from wild isolates collected from a forest near Toronto (see *Materials and Methods*). As expected, we assembled a genome consisting of six chromosomes, each of which is largely syntenic at a macro level with the genome assemblies from the other *Caenorhabditis* species. The difference in the genome lengths between *C. elegans* and *C. remanei* is quite large, from 102 Mb to well over 124 Mb. However, this degree of size variation appears to be typical for *Caenorhabditis* nematodes. For example, Stevens *et al.* (2019) showed that the genome sizes across the genus can vary from 65 to 140 Mb and that, overall, the size of the genome correlates with the number of genes but not necessarily the mode of reproduction.

This is the third *C. remanei* genome assembly generated by our group (PX439, PX356, and PX506; Table 1). The two previous chromosome-scale assemblies of other *C. remanei* strains (PX439 and PX356) were constructed with Illumina data and multiple mate-pair libraries (Fierst *et al.* 2015). However, the *C. remanei* genome has extended repetitive regions that failed to assemble using short reads. Further, strong segregation distortion among strains made it very difficult to construct the genetic map and definitively align shorter contigs to specific putative chromosomes. In this study, we used deep PacBio sequencing and Hi-C linkage information to overcome the repetitive regions and achieve better assembly characteristics. The combination of long-read and linkage data are a powerful toolset to produce chromosome-level assemblies, which are currently being increasingly used in a large number of species (*e.g.*, Gordon *et al.* 2016; Gong *et al.* 2018; VanBuren *et al.* 2018; Low *et al.* 2019).

In addition to genome assembly, we performed annotation of the new *C. remanei* reference genome, using full-length transcript data (Iso-Seq), which has proven to be an effective technique to create high-quality annotations (Gonzalez-Garay 2016), short-read transcriptome sequencing, protein sequences of related species, and a hybrid annotation pipeline. To validate predicted gene models, we additionally used the previous annotation of *C. remanei*, since it was manually curated and was mostly supported by RNA-seq data (Fierst *et al.* 2015). The genes that were not present in the previous annotation are supported by other lines of evidence, including genes predicted from the Iso-Seq data. We found a total of 26,308 genes in the *C. remanei* genome, a slight increase over previous estimates (Fierst *et al.* 2015) and reconfirmation that *C. remanei* appears to have more genes than the selfing species.

We compared the genomic organization of *C. remanei* and *C. elegans* using the latest available version of the VC2010 *C. elegans* genome, which is based on a modern strain derived from the classical N2 strain and which led to an enlargement of the N2-based genome by an additional 1.8 Mb of repetitive sequences (Yoshimura *et al.* 2019). *C. remanei* and *C. elegans* genomes are in high synteny in spite of multiple intrachromosomal rearrangements (Figure 1). We observed many more intra- than interchromosomal rearrangements, which is consistent with first comparative observations of the *C. elegans* and *C. briggsae* genomes, which saw a 10-fold difference in these rates (Stein *et al.* 2003). This overall pattern remains consistent even when comparing *C. elegans* to more distantly related genera of nematodes (Guiliano *et al.* 2002; Whitton *et al.* 2004; Mitreva *et al.* 2005).

One plausible explanation for this pattern is that the low rate of interchromosomal translocations is generated by the multilevel control of meiotic recombination in *Caenorhabditis*. Pairing of chromosomes during meiosis in *C. elegans* is initiated from specific regions ("pairing centers") located on the ends of homologous chromosomes (MacQueen *et al.* 2005; Tsai and McKee 2011), followed by chromosome synapsis via assembly of the synaptonemal complex along coupled chromosomes (MacQueen *et al.* 2002; Rog and Dernburg 2013). Crossovers in *C. elegans* can be formed only between properly synapsed regions (Lui and Colaiácovo 2013; Cahoon *et al.* 2019). Taken together, these molecular mechanisms permit meiotic recombination only between homologous regions linked *in cis* to the pairing centers, which presumably reduces
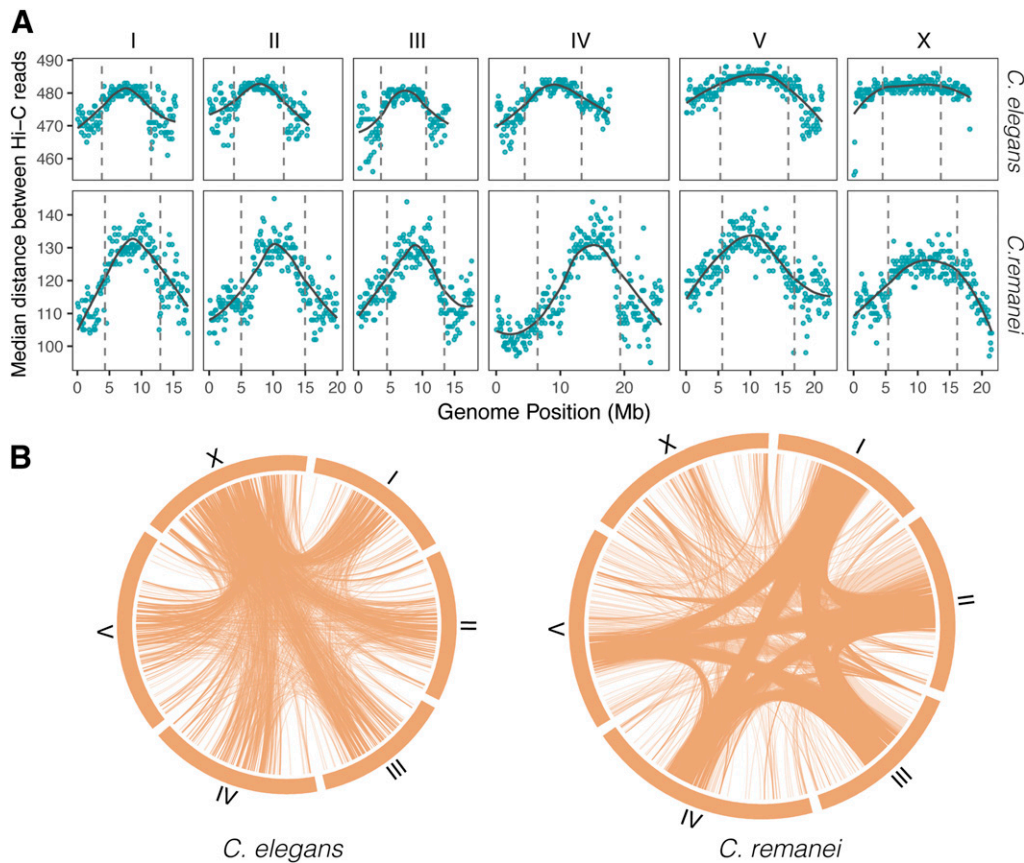
Figure 3 Genome landscape of median distances in Hi-C read pairs in *C. elegans* and *C. remanei* samples. (A) Distributions of distances between paired reads of *cis*-chromosome interactions. The vertical dashed lines show the boundaries of the central domain. Medians are estimated per 100-kb windows, with the gray lines representing the smoothed means of the values. The differences between values for *C. elegans* and *C. remanei* are due to the library size selection for Illumina sequencing in these two Hi-C experiments, and so only the relative patterns and not the absolute values are relevant here. (B) *Trans*-chromosome interactions in *C. elegans* and *C. remanei*. Lines represent contacts between 100-kb windows. For the *C. elegans* data set, only contacts with > 200 pairs of reads are shown; for the *C. remanei* data set, only contacts with > 100 pairs of reads are shown (the *C. remanei* Hi-C data set is two times smaller than the *C. elegans* data set). These filters emphasize differences in interaction density/location; the actual total number of interactions is approximately the same for all chromosomes.

the number of interchromosomal rearrangements, thereby resulting in the evolutionary stability of the nematode karyotype (Rog and Dernburg 2013).

The central domains of autosomes and a large portion of the *X* chromosome have more extended conservative regions between *C. remanei* and *C. elegans*. The similar pattern has been observed in comparative genomic studies of *C. elegans* and *C. briggsae* (Stein *et al.* 2003; Hillier *et al.* 2007). Apparently, the stability and conservation of the central regions is also connected to the recombinational landscape, as the central half chromosomes in *C. elegans* display a recombination rate that is several times lower than that observed on chromosome ends (Rockman and Kruglyak 2009), *C. briggsae* (Ross *et al.* 2011), as well as in *C. remanei* (A. A. Teterina, J. H. Willis, P. C. Phillips personal communication), without definitive hotspots of recombination (Kaur and Rockman 2014). Variation in recombination rate on the *X* chromosome is less than that on autosomes (Bernstein and Rockman 2016) and, because of the *XX/X*0 sex determination system of nematodes, the population size of the sex chromosome is three-quarters that of the autosomes (Wright 1931). So, orthologs of *C. elegans* and *C. remanei* located on the *X* chromosome are more conserved on average, likely because selection against deleterious mutations on the sex chromosome

is greater than on autosomes (Montgomery *et al.* 1987; Coghlan and Wolfe 2002).

The chromosomes of *C. elegans* (*C. elegans* Sequencing Consortium 1998) and *C. remanei* also have a very similar pattern of gene organization, with a central region (the central domain or "central gene cluster") (Barnes *et al.* 1995) characterized by high gene density, shorter genes and introns, lower GC content (Figure S5 and Table S4), and almost two times lower abundance of repetitive elements compared to chromosome arms. Repetitive elements in *C. elegans* and *C. remanei* are more abundant in the peripheries of chromosomes and, respectively, leave less room for protein-coding genes in those regions. About 28% of the total intron lengths in these nematodes are occupied by transposable elements, which could partially explain the increase of the gene lengths and intron fractions on the arms. The positive correlation of intron size with recombination rate and transposable elements has been previously observed in *C. elegans* (Prachumwat *et al.* 2004; Li *et al.* 2009b). The central gene clusters and transposable elements enriched in the arms are common, and are likely the ancestral pattern observed in *C. elegans*, *C. briggsae*, *C. tropicalis*, and *C. remanei*, yet distinct in *C. inopinata* (Woodruff and Teterina, 2019 *preprint*).

Use of Hi-C data in the genome assembly allows us to perform a preliminary analysis of the 3D chromatin organization across mixed developmental stages in *C. elegans* and *C. remanei*. The central domains show more *cis*-chromosome interactions than the peripheral parts of chromosomes in *C. remanei* (Figure 3). In *C. elegans*, variation in interaction intensity across the chromosome is somewhat less perceptible, probably because of minor differences in the fractions of genes on the central domains *vs.* arms. In both species, central regions show more distant interactions than arms. All chromosomes have numerous *trans*-chromosome interactions that are more tightly localized in the central regions. This pattern can be explained both by the densities of genes in the central domains and by technical issues with mapping of the reads to the repetitive regions. In contrast to the autosomes, the pattern of *trans*-chromosome activity is more dissimilar in *C. elegans* and *C. remanei*. This could be caused by species-specific differences or by the fact that the developmental stages of the samples do not strictly correspond for the two species (the *C. elegans* data set used early embryos whereas the *C. remanei* sample included all stages of the life cycle). In this case, both *X* chromosomes are active in hermaphrodites (*XX*), but their activity is reduced by one-half by a dosage-compensation mechanism in all tissues in *C. elegans* after the 30-cell stage (gastrulation) (Meyer 2005; Strome *et al.* 2014; Crane *et al.* 2015; Brejc *et al.* 2017). The presences of individuals at the early developmental stages could therefore potentially affect the extents of interactions with *X* chromosome observed within the *C. elegans* sample. Dosage compensation suppresses gene expression on both *X* chromosomes, modulates chromatin conformation by forming topologically associated domains, and partially compresses both *X* chromosomes (Meyer 2010; Lau *et al.* 2014; Brejc *et al.* 2017). All of these structural changes could potentially affect the relative intensities and availabilities of interactions between the *X* chromosome and autosomes.

What might drive these interchromosomal interactions? *Cis*- and *trans*-chromosome interactions could mediate transcriptional activity through colocalization of transcriptional factors on gene regulatory regions (Miele and Dekker 2008; Pai and Engelke 2010; Maass *et al.* 2019). Genome activity and the spatial organization of a genome are dynamic properties, and chromatin accessibility in *C. elegans* is tissue-specific, changing over developmental time (Daugherty *et al.* 2017; Jänes *et al.* 2018). However, *C. elegans* tends to have active euchromatin in the central parts of chromosomes and silent heterochromatin in the arms, which are anchored to the nuclear membrane (Ikegami *et al.* 2010; Liu *et al.* 2011; Mattout *et al.* 2015; Solovei *et al.* 2016; Cabianca *et al.* 2019). This pattern of regulation is consistent with the pattern of interactions that we observe. Nevertheless, much more work needs to be conducted, particularly aimed at stage- and tissue-specific effects, before the role and dynamics of spatial chromosome interaction in *Caenorhabditis* can be fully revealed.

Overall, despite numerous within-chromosome rearrangements, *C. elegans* and *C. remanei* show similar patterns of chromosomal structure and activity. The chromosome-level assembly of *C. remanei* presented here provides a solid new platform for experimental evolution, comparative and population genomics, and the study of genome function and architecture.

Acknowledgments

Literature Cited

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. 215: 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anders, S., P. T. Pyl, and W. Huber, 2015 HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31: 166–169. https://doi.org/10.1093/bioinformatics/btu638

Bao, W., K. K. Kojima, and O. Kohany, 2015 Repbase update, a database of repetitive elements in eukaryotic genomes. Mob. DNA 6: 11. https://doi.org/10.1186/s13100-015-0041-9

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

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman *et al.*, 2012 GenBank. Nucleic Acids Res. 41: D36–D42. https://doi.org/10.1093/nar/gks1195

Bernstein, M. R., and M. V. Rockman, 2016 Fine-scale crossover rate variation on the Caenorhabditis elegans X chromosome. G3 (Bethesda) 6: 1767–1776.

Blaxter, M., 1998 *Caenorhabditis elegans* is a nematode. Science 282: 2041–2046. https://doi.org/10.1126/science.282.5396.2041

Brejc, K., Q. Bian, S. Uzawa, B. S. Wheeler, E. C. Anderson *et al.*, 2017 Dynamic control of X chromosome conformation and repression by a histone H4K20 demethylase. Cell 171: 85–102.e23. https://doi.org/10.1016/j.cell.2017.07.041

Brenner, S., 1973 The genetics of behaviour. Br. Med. Bull. 29: 269–271. https://doi.org/10.1093/oxfordjournals.bmb.a071019

Brenner, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics 77: 71–94.

Cabianca, D. S., C. Muñoz-Jiménez, V. Kalck, D. Gaidatzis, J. Padeken *et al.*, 2019 Active chromatin marks drive spatial sequestration of heterochromatin in *C. elegans* nuclei. Nature 569: 734–739. https://doi.org/10.1038/s41586-019-1243-y

Cahoon, C. K., J. M. Helm, and D. E. Libuda, 2019 Synaptonemal complex central region proteins promote localization of procrossover factors to recombination events during *Caenorhabditis elegans* meiosis. Genetics 213: 395–409. https://doi.org/10.1534/genetics.119.302625

Castillo, D. M., M. K. Burger, C. M. Lively, and L. F. Delph, 2015 Experimental evolution: assortative mating and sexual selection, independent of local adaptation, lead to reproductive isolation in the nematode *Caenorhabditis remanei*. Evolution 69: 3141–3155. https://doi.org/10.1111/evo.12815

C. elegans Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282: 2012–2018 [corrigenda: Science 283: 35 (1999)];

[corrigenda: Science 283: 2103 (1999)]; [corrigenda: Science 285: 1493 (1999)].

Chin, C.-S., P. Peluso, F. J. Sedlazeck, M. Nattestad, G. T. Concepcion et al., 2016 Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods 13: 1050–1054. https://doi.org/10.1038/nmeth.4035

Coghlan, A., and K. H. Wolfe, 2002 Fourfold faster rate of genome rearrangement in nematodes than in Drosophila. Genome Res. 12: 857–867. https://doi.org/10.1101/gr.172702

Coghlan, A., I. J. Tsai, and M. Berriman, 2018 Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. Protoc. Exch. DOI: 10.1038/protex.2018.054. https://doi.org/10.1038/protex.2018.054

Crane, E., Q. Bian, R. P. McCord, B. R. Lajoie, B. S. Wheeler et al., 2015 Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature 523: 240–244. https://doi.org/10.1038/nature14450

Cutter, A. D., and B. Charlesworth, 2006 Selection intensity on preferred codons correlates with overall codon usage bias in Caenorhabditis remanei. Curr. Biol. 16: 2053–2057. https://doi.org/10.1016/j.cub.2006.08.067

Cutter, A. D., S. E. Baird, and D. Charlesworth, 2006 High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of Caenorhabditis remanei. Genetics 174: 901–913. https://doi.org/10.1534/genetics.106.061879

Daugherty, A. C., R. W. Yeo, J. D. Buenrostro, W. J. Greenleaf, A. Kundaje et al., 2017 Chromatin accessibility dynamics reveal novel functional enhancers in C. elegans. Genome Res. 27: 2096–2107. https://doi.org/10.1101/gr.226233.117

Dey, A., Y. Jeon, G.-X. Wang, and A. D. Cutter, 2012 Global population genetic structure of Caenorhabditis remanei reveals incipient speciation. Genetics 191: 1257–1269. https://doi.org/10.1534/genetics.112.140418

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski et al., 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dolgin, E. S., B. Charlesworth, S. E. Baird, and A. D. Cutter, 2007 Inbreeding and outbreeding depression in Caenorhabditis nematodes. Evolution 61: 1339–1352. https://doi.org/10.1111/j.1558-5646.2007.00118.x

Dunham, A., L. Matthews, J. Burton, J. Ashurst, K. Howe et al., 2004 The DNA sequence and analysis of human chromosome 13. Nature 428: 522–528. https://doi.org/10.1038/nature02379

Edgar, R. C., 2010 Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26: 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Emms, D. M., and S. Kelly, 2019 OrthoFinder2: fast and accurate phylogenomic orthology analysis from gene sequences. Genome Biol. 20: 238. https://doi.org/10.1186/s13059-019-1832-y

Fierst, J. L., J. H. Willis, C. G. Thomas, W. Wang, R. M. Reynolds et al., 2015 Reproductive mode and the evolution of genome size and structure in Caenorhabditis nematodes. PLoS Genet. 11: e1005323 (erratum PLoS Genet. 11: e1005497). https://doi.org/10.1371/journal.pgen.1005323

Finn, R. D., P. Coggill, R. Y. Eberhardt, S. R. Eddy, J. Mistry et al., 2015 The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44: D279–D285. https://doi.org/10.1093/nar/gkv1344

Gerstein, M. B., Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff et al., 2010 Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science 330: 1775–1787. https://doi.org/10.1126/science.1196914

Gong, G., C. Dan, S. Xiao, W. Guo, P. Huang et al., 2018 Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. Gigascience 7: giy120.

Gonzalez-Garay, M. L., 2016 Introduction to isoform sequencing using Pacific Biosciences technology (Iso-Seq), pp. 141–160 in Transcriptomics and Gene Regulation. Springer, Dordrecht, The Netherlands.

Gordon, D., J. Huddleston, M. J. Chaisson, C. M. Hill, Z. N. Kronenberg et al., 2016 Long-read sequence assembly of the gorilla genome. Science 352: aae0344. https://doi.org/10.1126/science.aae0344

Graustein, A., J. M. Gaspar, J. R. Walters, and M. F. Palopoli, 2002 Levels of DNA polymorphism vary with mating system in the nematode genus Caenorhabditis. Genetics 161: 99–107.

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

Guiliano, D., N. Hall, S. Jones, L. Clark, C. Corton et al., 2002 Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. Genome Biol. 3: RESEARCH0057.

Haas, B. J., S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen et al., 2008 Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9: R7. https://doi.org/10.1186/gb-2008-9-1-r7

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood et al., 2013 De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8: 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hillier, L. W., R. D. Miller, S. E. Baird, A. Chinwalla, L. A. Fulton et al., 2007 Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 5: e167. https://doi.org/10.1371/journal.pbio.0050167

Holt, C., and M. Yandell, 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 491. https://doi.org/10.1186/1471-2105-12-491

Hubley, R., R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones et al., 2015 The Dfam database of repetitive DNA families. Nucleic Acids Res. 44: D81–D89. https://doi.org/10.1093/nar/gkv1272

Ikegami, K., T. A. Egelhofer, S. Strome, and J. D. Lieb, 2010 Caenorhabditis elegans chromosome arms are anchored to the nuclear membrane via discontinuous association with LEM-2. Genome Biol. 11: R120. https://doi.org/10.1186/gb-2010-11-12-r120

Jänes, J., Y. Dong, M. Schoof, J. Serizay, A. Appert et al., 2018 Chromatin accessibility dynamics across C. elegans development and ageing. Elife 7: e37344. https://doi.org/10.7554/eLife.37344

Jovelin, R., B. Ajie, and P. Phillips, 2003 Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus Caenorhabditis. Mol. Ecol. 12: 1325–1337. https://doi.org/10.1046/j.1365-294X.2003.01805.x

Jovelin, R., J. P. Dunham, F. S. Sung, and P. C. Phillips, 2009 High nucleotide divergence in developmental regulatory genes contrasts with the structural elements of olfactory pathways in Caenorhabditis. Genetics 181: 1387–1397. https://doi.org/10.1534/genetics.107.082651

Kanzaki, N., I. J. Tsai, R. Tanaka, V. L. Hunt, D. Liu et al., 2018 Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. Nat. Commun. 9: 3216. https://doi.org/10.1038/s41467-018-05712-5

Kaur, T., and M. V. Rockman, 2014 Crossover heterogeneity in the absence of hotspots in Caenorhabditis elegans. Genetics 196: 137–148. https://doi.org/10.1534/genetics.113.158857

Kiontke, K. C., M.-A. Félix, M. Ailion, M. V. Rockman, C. Braendle *et al.*, 2011 A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. BMC Evol. Biol. 11: 339. https://doi.org/10.1186/1471-2148-11-339

Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway *et al.*, 2004 Versatile and open software for comparing large genomes. Genome Biol. 5: R12. https://doi.org/10.1186/gb-2004-5-2-r12

Lau, A. C., K. Nabeshima, and G. Csankovszki, 2014 The *C. elegans* dosage compensation complex mediates interphase X chromosome compaction. Epigenetics Chromatin 7: 31. https://doi.org/10.1186/1756-8935-7-31

Li, H., 2011 A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27: 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009a The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li, H., G. Liu, and X. Xia, 2009b Correlations between recombination rate and intron distributions along chromosomes of *C. elegans*. Prog. Nat. Sci. 19: 517–522. https://doi.org/10.1016/j.pnsc.2008.06.019

Liu, T., A. Rechtsteiner, T. A. Egelhofer, A. Vielle, I. Latorre *et al.*, 2011 Broad chromosomal domains of histone modification patterns in *C. elegans*. Genome Res. 21: 227–236. https://doi.org/10.1101/gr.115519.110

Llorens, C., R. Futami, L. Covelli, L. Domínguez-Escribá, J. M. Viu *et al.*, 2010 The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 39: D70–D74. https://doi.org/10.1093/nar/gkq1061

Low, W. Y., R. Tearle, D. M. Bickhart, B. D. Rosen, S. B. Kingan *et al.*, 2019 Chromosome-level assembly of the water buffalo genome surpasses human and goat genomes in sequence contiguity. Nat. Commun. 10: 260. https://doi.org/10.1038/s41467-018-08260-0

Lowe, T. M., and S. R. Eddy, 1997 tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964. https://doi.org/10.1093/nar/25.5.955

Lui, D. Y., and M. P. Colaiácovo, 2013 Meiotic development in *Caenorhabditis elegans*, pp. 133–170 in *Germ Cell Development in C. elegans*. Springer-Verlag, New York.

Maass, P. G., A. R. Barutcu, and J. L. Rinn, 2019 Interchromosomal interactions: a genomic love story of kissing chromosomes. J. Cell Biol. 218: 27–38. https://doi.org/10.1083/jcb.201806052

MacQueen, A. J., M. P. Colaiácovo, K. McDonald, and A. M. Villeneuve, 2002 Synapsis-dependent and -independent mechanisms stabilize homolog pairing during meiotic prophase in *C. elegans*. Genes Dev. 16: 2428–2442. https://doi.org/10.1101/gad.1011602

MacQueen, A. J., C. M. Phillips, N. Bhalla, P. Weiser, A. M. Villeneuve *et al.*, 2005 Chromosome sites play dual roles to establish homologous synapsis during meiosis in *C. elegans*. Cell 123: 1037–1050. https://doi.org/10.1016/j.cell.2005.09.034

Mattout, A., D. S. Cabianca, and S. M. Gasser, 2015 Chromatin states and nuclear organization in development—a view from the nuclear lamina. Genome Biol. 16: 174. https://doi.org/10.1186/s13059-015-0747-5

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303. https://doi.org/10.1101/gr.107524.110

Meyer, B. J., 2005 X-chromosome dosage compensation (June 25, 2005), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.8.1, http://www.wormbook.org.

Meyer, B. J., 2010 Targeting X chromosomes for repression. Curr. Opin. Genet. Dev. 20: 179–189. https://doi.org/10.1016/j.gde.2010.03.008

Miele, A., and J. Dekker, 2008 Long-range chromosomal interactions and gene regulation. Mol. Biosyst. 4: 1046–1057. https://doi.org/10.1039/b803580f

Mitreva, M., M. L. Blaxter, D. M. Bird, and J. P. McCarter, 2005 Comparative genomics of nematodes. Trends Genet. 21: 573–581. https://doi.org/10.1016/j.tig.2005.08.003

Montgomery, E., B. Charlesworth, and C. H. Langley, 1987 A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. Genet. Res. 49: 31–41. https://doi.org/10.1017/S0016672300026707

Navarro, D., 2013 *Learning Statistics with R: A Tutorial for Psychology Students and Other Beginners: Version 0.5*. Available at: https://open.umn.edu/opentextbooks/textbooks/learning-statistics-with-r-a-tutorial-for-psychology-students-and-other-beginners. Accessed: March 2020.

Pai, D. A., and D. R. Engelke, 2010 Spatial organization of genes as a component of regulated expression. Chromosoma 119: 13–25. https://doi.org/10.1007/s00412-009-0236-2

Pires-daSilva, A., 2007 Evolution of the control of sexual identity in nematodes. *Semin Cell Dev. Biol.* 18: 362–370.

Prachumwat, A., L. DeVincentis, and M. F. Palopoli, 2004 Intron size correlates positively with recombination rate in *Caenorhabditis elegans*. Genetics 166: 1585–1590. https://doi.org/10.1534/genetics.166.3.1585

Putnam, N. H., B. L. O'Connell, J. C. Stites, B. J. Rice, M. Blanchette *et al.*, 2016 Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 26: 342–350. https://doi.org/10.1101/gr.193474.115

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder *et al.*, 2005 InterProScan: protein domains identifier. Nucleic Acids Res. 33: W116–W120. https://doi.org/10.1093/nar/gki442

Quinlan, A. R., and I. M. Hall, 2010 BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26: 841–842. https://doi.org/10.1093/bioinformatics/btq033

R Core Team, 2018 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

Reynolds, R. M., and P. C. Phillips, 2013 Natural variation for lifespan and stress response in the nematode *Caenorhabditis remanei*. PLoS One 8: e58212. https://doi.org/10.1371/journal.pone.0058212

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

Rog, O., and A. F. Dernburg, 2013 Chromosome pairing and synapsis during *Caenorhabditis elegans* meiosis. Curr. Opin. Cell Biol. 25: 349–356. https://doi.org/10.1016/j.ceb.2013.03.003

Ross, J. A., D. C. Koboldt, J. E. Staisch, H. M. Chamberlin, B. P. Gupta *et al.*, 2011 *Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. 7: e1002174. https://doi.org/10.1371/journal.pgen.1002174

Sikkink, K. L., C. M. Ituarte, R. M. Reynolds, W. A. Cresko, and P. C. Phillips, 2014a The transgenerational effects of heat stress in the nematode *Caenorhabditis remanei* are negative and rapidly eliminated under direct selection for increased stress resistance in larvae. Genomics 104: 438–446. https://doi.org/10.1016/j.ygeno.2014.09.014

Sikkink, K. L., R. M. Reynolds, C. M. Ituarte, W. A. Cresko, and P. C. Phillips, 2014b Rapid evolution of phenotypic plasticity and shifting thresholds of genetic assimilation in the nematode Caenorhabditis remanei. G3 (Bethesda) 4: 1103–1112.

Sikkink, K. L., R. M. Reynolds, W. A. Cresko, and P. C. Phillips, 2015 Environmentally induced changes in correlated responses to selection reveal variable pleiotropy across a complex genetic network. Evolution 69: 1128–1142. https://doi.org/10.1111/evo.12651

Sikkink, K. L., R. M. Reynolds, C. M. Ituarte, W. A. Cresko, and P. C. Phillips, 2019 Environmental and evolutionary drivers of the modular gene regulatory network underlying phenotypic plasticity for stress resistance in the nematode *Caenorhabditis remanei*. G3 (Bethesda) 9: 969–982.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov, 2015 BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit, A. F., and R. Hubley, 2008 RepeatModeler Open-1.0. Available at: http://www.repeatmasker.org. Accessed: March 2020.

Smit, A. F., R. Hubley, and P. Green, 2015 RepeatMasker Open-4.0. 2013–2015. Available at: http://www.repeatmasker.org. Accessed: March 2020.

Solovei, I., K. Thanisch, and Y. Feodorova, 2016 How to rule the nucleus: divide et impera. Current Op. Cell Biol. (Henderson NV) 40: 47–59.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack et al., 2006 AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 34: W435–W439. https://doi.org/10.1093/nar/gkl200

Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent et al., 2003 The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 1: e45. https://doi.org/10.1371/journal.pbio.0000045

Steinbiss, S., U. Willhoeft, G. Gremme, and S. Kurtz, 2009 Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. 37: 7002–7013. https://doi.org/10.1093/nar/gkp759

Stevens, L., M. A. Félix, T. Beltran, C. Braendle, C. Caurcel et al., 2019 Comparative genomics of 10 new Caenorhabditis species. Evol. Lett. 3: 217–236. https://doi.org/10.1002/evl3.110

Strome, S., W. G. Kelly, S. Ercan, and J. D. Lieb, 2014 Regulation of the X chromosomes in Caenorhabditis elegans. Cold Spring Harb. Perspect. Biol. 6: a018366. https://doi.org/10.1101/cshperspect.a018366

Ter-Hovhannisyan, V., A. Lomsadze, Y. O. Chernoff, and M. Borodovsky, 2008 Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. Genome Res. 18: 1979–1990. https://doi.org/10.1101/gr.081612.108

Thorvaldsdóttir, H., J. T. Robinson, and J. P. Mesirov, 2013 Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14: 178–192. https://doi.org/10.1093/bib/bbs017

Tsai, J.-H., and B. D. McKee, 2011 Homologous pairing and the role of pairing centers in meiosis. J. Cell Sci. 124: 1955–1963. https://doi.org/10.1242/jcs.006387

VanBuren, R., C. M. Wai, M. Colle, J. Wang, S. Sullivan et al., 2018 A near complete, chromosome-scale assembly of the black raspberry (Rubus occidentalis) genome. Gigascience 7: giy094. https://doi.org/10.1093/gigascience/giy094

Waterston, R., C. Martin, M. Craxton, C. Huynh, A. Coulson et al., 1992 A survey of expressed genes in Caenorhabditis elegans. Nat. Genet. 1: 114–123. https://doi.org/10.1038/ng0592-114

Whitton, C., J. Daub, M. Quail, N. Hall, J. Foster et al., 2004 A genome sequence survey of the filarial nematode Brugia malayi: repeats, gene discovery, and comparative genomics. Mol. Biochem. Parasitol. 137: 215–227. https://doi.org/10.1016/j.mol-biopara.2004.05.013

Wickham, H., 2016 *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Woodruff, G. C., and A. A. Teterina, 2019 Degradation of the repetitive genomic landscape in a close relative of C. elegans. bioRxiv. (Preprint posted October 8, 2019).https://doi.org/10.1101/797035

Wright, S., 1931 Evolution in Mendelian populations. Genetics 16: 97–159.

Wu, T. D., and C. K. Watanabe, 2005 GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics 21: 1859–1875. https://doi.org/10.1093/bioinformatics/bti310

Ye, C., G. Ji, and C. Liang, 2016 detectMITE: a novel approach to detect miniature inverted repeat transposable elements in genomes. Sci. Rep. 6: 19688. https://doi.org/10.1038/srep19688

Yin, D., E. M. Schwarz, C. G. Thomas, R. L. Felde, I. F. Korf et al., 2018 Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science 359: 55–61. https://doi.org/10.1126/science.aao0827

Yoshimura, J., K. Ichikawa, M. J. Shoura, K. L. Artiles, I. Gabdank et al., 2019 Recompleting the Caenorhabditis elegans genome. Genome Res. 29: 1009–1022. https://doi.org/10.1101/gr.244830.118

Communicating editor: V. Reinke