SMC-mediated dosage compensation in *C. elegans* evolved in the presence of an ancestral nematode mechanism

Avrami Aharonoff*, Jun Kim, Aaliyah Washington, Sevinç Ercan*

Department of Biology, Center for Genomics and Systems Biology, New York University, New York, NY 10003

*Correspondence: aaharonoff@nyu.edu, se71@nyu.edu

Key words: dosage compensation, sex chromosomes, nematodes, pristionchus, caenorhabditis, oscheius, Hi-C, TAD, condensin, X chromosome, Structural Maintenance of Chromosomes, SMC complexes, evolution, transcription, H4K20me1

ABSTRACT

Mechanisms of X chromosome dosage compensation have been studied extensively in a few model species representing clades of shared sex chromosome ancestry. However, the diversity within each clade as a function of sex chromosome evolution is largely unknown. Here, we anchor ourselves to the nematode *Caenorhabditis elegans*, for which a well-studied mechanism of dosage compensation occurs through a specialized structural maintenance of chromosomes (SMC) complex, and explore the diversity of dosage compensation in the surrounding phylogeny of nematodes. Through phylogenetic analysis of the *C. elegan*s dosage compensation complex and a survey of its epigenetic signatures, including X-specific topologically associating domains (TADs) and X-enrichment of H4K20me1, we found that the condensin-mediated mechanism evolved recently in the lineage leading to *Caenorhabditis* through an SMC-4 duplication. Intriguingly, an independent duplication of SMC-4 and the presence of X-specific TADs in *Pristionchus pacificus* suggest that condensin-mediated dosage compensation arose more than once. mRNA-seq analyses of gene expression in several nematode species indicate that dosage compensation itself is ancestral, as expected from the ancient XO sex determination system. Indicative of the ancestral mechanism, H4K20me1 is enriched on the X chromosomes in *Oscheius tipulae*, which does not contain X-specific TADs or SMC-4 paralogs. Together, our results indicate that the dosage compensation system in *C. elegans* is surprisingly new, and condensin may have been co-opted repeatedly in nematodes, suggesting that the process of evolving a chromosome-wide gene regulatory mechanism for dosage compensation is constrained.

Significance statement

X chromosome dosage compensation mechanisms evolved in response to Y chromosome degeneration during sex chromosome evolution. However, establishment of dosage compensation is not an endpoint. As sex chromosomes change, dosage compensation strategies may have also changed. In this study, we performed phylogenetic and epigenomic analyses surrounding *Caenorhabditis elegans* and found that the condensin-mediated dosage compensation mechanism in *C. elegans* is surprisingly new, and has evolved in the presence of an ancestral mechanism. Intriguingly, condensin-based dosage compensation may have evolved more than once in the nematode lineage, the other time in *Pristionchus*. Together, our work highlights a previously unappreciated diversity of dosage compensation mechanisms within a clade, and suggests constraints in evolving new mechanisms in the presence of an existing one.

INTRODUCTION

According to the current model of sex chromosome evolution, heteromorphic X and Y chromosomes descend from a homologous, recombining set of autosomes that acquire a sex determining role[1–5]. Suppression of recombination between these newly formed sex chromosomes leads to the degeneration of the Y chromosome and almost complete hemizygosity of the male X chromosome[1,6,7]. Maintenance of the ancestral gene dose, however, is particularly important for dosage sensitive gene networks with a mix of autosomal and sex chromosomal genes[8]. In order to maintain ancestral gene dose, mechanisms of sex chromosome specific gene regulation are selected for. These genetic and epigenetic mechanisms that compensate for the imbalance in gene dose between males and females are collectively referred to as “dosage compensation”[3].

Sex chromosomes have co-evolved with dosage compensation many times. Therefore, clades defined by a shared ancestral X chromosome can also be defined by their dosage compensation mechanism. For example, mammals, flies, and nematodes all use different strategies. In *Mus musculus* (mouse), a long noncoding RNA, *Xist*, initiates several heterochromatic pathways to randomly inactivate one of the X chromosomes in females (XX)[9–11]. In *Drosophila melanogaster* (fly), a histone acetylating ribonucleoprotein complex, *male specific lethal* (MSL) activates transcription by about two-fold on the single X chromosome in males (XY)[12,13]. Finally, in the nematode *Caenorhabditis elegans*, the *dosage compensation complex* (DCC) is driven by an X-specific Structural Maintenance of Chromosome (SMC) complex that represses transcription by about two-fold in hermaphrodites (XX)[14–18] (**Figure 1A**).

The descent of heteromorphic X and Y chromosomes is *not* an end point. Chromosomal duplications, deletions, fusions, and translocations continue to change sex chromosomes[19–23]. Therefore, although initial dosage compensation mechanisms evolve in response to Y chromosome degradation, they continue to evolve in response to X chromosome evolution. For example, in two species of *Drosophila*, a translocation of autosomal sequences to the X chromosome selected for DNA sequence motifs capable of recruiting the MSL complex to the neo-X region[24–26]. While the motifs required for MSL recruitment remained similar in the *Drosophila* system, in two *Caenorhabditis* species, *C. elegans* and *C. briggsae* (estimated to have diverged from *C. elegans* 15-30 million years ago), the DNA sequence motifs that recruit the DCC have functionally diverged[27].

The studies performed in *Drosophila* used species in which MSL complex subunits were readily identified by sequence homology[24–26]. Similarly, subunits of the *C. elegans* DCC were identified by sequence homology in *C. briggsae*[27]. Studying the mechanism of dosage compensation in more distantly related species, however, is difficult due to the rapid evolution of proteins involved in sex determination and dosage compensation[2,21]. Furthermore, until recently, chromosome scale assemblies in non-model organisms were few and far between. Therefore, the field remains largely ignorant to the diversity in dosage compensation mechanisms in species that have diverged significantly, but share an X chromosome ancestor.

Here, we used a combination of phylogenetic and genomic approaches to explore dosage compensation in several species of nematodes with significant differences in X chromosome content. We anchored our work in *C. elegans*, a model organism in which XX and XO animals develop as hermaphrodites and males, respectively[15]. The core of the *C. elegans* DCC is a condensin complex belonging to the SMC family[17,28–31]. Most animals harbor two canonical condensins (I and II) that mediate DNA loops to compact chromosomes, and are essential for chromosome segregation in mitosis and meiosis[29]. Condensins I and II share a heterodimer of SMC-2 and SMC-4 subunits, which interact with a different set of three chromosome-associated proteins (CAP)[32]. In *C. elegans*, the third condensin (I^{DC}) is formed by replacing SMC-4 with DPY-27 in condensin I (**Figure 1A**)[17]. Condensin I^{DC} (hereafter I-DC) interacts with five non-condensin proteins to form the DCC[33].

The two major epigenetic changes mediated by the DCC on the X chromosomes are (1) the formation of loop-anchored topologically associating domains (TADs) and (2) the enrichment of H4K20me1 in females and hermaphrodites[34–37]. To explore dosage compensation across distantly related nematodes, we first generated a phylogeny of SMC-4 and DPY-27 in Rhabditina (clade V), the clade representing *C. elegans*. We then looked for condensin-mediated epigenetic phenotypes by performing Hi-C and ChIP-seq in multiple species. Based on the presence of two SMC-4 paralogs, X-specific loop anchored TADs, and X enrichment of H4K20me1, we conclude that condensin-mediated dosage compensation arose more than once, the other time being in *Pristionchus.* In *Oscheius*, we found evidence for dosage compensation and upregulation of H4K20me1 without TADs, suggesting that the ancestral form of regulation may be through the well conserved histone demethylase DPY-21, a subunit of the *C. elegans* DCC. We propose that condensin I-DC arose in the lineage leading to *Caenorhabditis* in the presence of an ancestral compensation mechanism. Our findings highlight the continued and constrained evolution of dosage compensation mechanisms, and a previously underappreciated diversity of strategies in species with a shared X chromosome ancestry.

MATERIALS AND METHODS

Phylogenetic analysis

Bioproject numbers can be found in **supplemental file 1**. To assess annotation quality, BUSCO scores were called in protein mode against the nematoda lineage. We used version 3.0.2 (OBD9 dataset) for the diplogastrids, and version 5.3.0 (ODB10 dataset) for the rest[38–41]. We called orthologs using Orthofinder (version 2.5.4) by first extracting the longest transcript per gene with the primary_transcripts.py script, and then running Orthofinder with the default parameters[42,43]. We removed orthologs that were less than half the size of SMC-4 in *C. elegans*. These sequences represent either partial duplications or duplications followed by deletions and are unlikely to have the function of SMC-4 preserved. Furthermore, the resulting gaps in the alignment would affect downstream analysis. We aligned orthologs using MAFFT (version 7.475) with the following parameters: --local pair --maxiterate 1000[44]. We generated a maximum-likelihood gene tree using IQ-Tree (version 2.2.0) on the aligned sequences with the

following parameters: --seqtype AA -m MFP -B 1000[45–47]. The trees were edited in FigTree (version 1.4.4) and TreeViewer[48].

Strains and growth conditions

We used *Pristionchus pacificus* strain PS312, and *Oscheius tipulae* strain CEW1. Both were maintained on NGM plates (3 g NaCl, 2.5 g peptone, 20 g agar, 1 mM calcium chloride, 5 mg/L cholesterol, 1 mM magnesium sulfate, and 25 mM potassium phosphate buffer), seeded with the streptomycin-resistant *Escherichia coli* strain OP-50, and fed as needed with a concentrated HB101 strain. *P. pacificus* was maintained at 20 ˚C, and *O. tipulae* at 22 ˚C.

Hi-C

P. pacificus and *O. tipulae* were propagated for collection on NGM plates supplemented with agarose (10 g agar, 7 g agarose) to prevent burrowing. *P. pacificu*s was filtered through a 20 µ nylon mesh to separate J2 larvae (flow through), which were fixed with 2% formaldehyde for 30 minutes, washed with M9, and stored at -80 ˚C. *O. tipulae* was washed off plates with sterile M9 and a cell spreader to unstick the embryos from the plate. Adults and larvae were killed using a bleach solution (0.5 M NaOH, 1% sodium hypochlorite) for 2 minutes, washed with M9, and nutated in M9 for 24 hours at 22 ˚C to arrest worms at the L1 stage. Prior to plating, *O. tipulae* worms were filtered through a 15 µ nylon mesh filter to separate the hatched L1 larvae from carcasses and unhatched embryos. L1s were grown for 24 hours at 22 ˚C, fixed with 2% formaldehyde for 30 minutes, washed with M9, and stored at -80 ˚C. Prior to fixing, the extent of germline proliferation in *O. tipulae* was determined by DIC microscopy.

Larvae were removed from -80 ˚C and resuspended in PBS with 1 mM PMSF. A mortar and pestle (BioSpec Products, catalog # 206) was placed on dry ice and filled with liquid nitrogen. Drops of larvae were pipetted into the mortar, and pulverized into a fine powder while submerged in liquid nitrogen. The powdered larvae were crosslinked as described in the Arima Hi-C+ High Coverage Kit (catalog # A101030). Hi-C (restriction enzymes: DpnII, HinfI, DdeI, and MseI) was then performed as per Arima's instructions. Library preparation was performed with the KAPA Hyper Prep Kit (KK8502) as per Arima's instructions. Paired-end (100 bp) sequencing was performed at the Genomics Core at the Center for Systems and Genome Biology, New York University using the Illumina NovaSeq 6000 platform.

Re-scaffolding and liftover

Hi-C fastq files were mapped to the *P. pacificus* genome (PRJNA12644), and biological replicates were merged using Arima Genomics' mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline). The output bam file was used to re-scaffold the *P. pacificus* genome with YaHS (version 1.1)[49]. The following parameters were used: -r 1000, 2000, 5000, 10000, 20000, 50000, 100000. The draft genome was curated with Juicebox Assembly Tools (JBAT) in Juicebox (see code availability)[50]. Liftoff (version 1.6.3) was used to liftover the protein coding genes to the updated assembly with the default parameters.[51]

Hi-C data processing and analysis

Hic files were generated using Juicer (version 1.5.7)[50,52]. Biological replicates were merged using Juicer's mega.sh script. All output hic files (minimum mapping quality of 30) were converted to cool files using the hicConvertFormat tool in HiCExplorer (version 3.6) with the following parameters: --inputFormat hic --outputFormat cool. For observed/expected plots, we used the hicTransform tool in HiCExplorer with the following parameters: --method obs_exp_lieberman before continuing. Count data was loaded from the cool file using the following parameters: -- inputFormat cool --outputFormat cool --load_raw_values. This raw coo/ file was then balanced using Cooler (version 0.8.11) with the following parameters: --max-iters 500 --mad-max 5 --ignore-diags 2. The balanced cool file was used in Cooltools (version 0.4.0) to generate the Hi-C matrixes, and compute the P(s), its derivative, and insulation scores[53]. All plots were made with Matplotlib (version 3.4.3)[54].

ChIP-seq

Larvae were collected as they were for Hi-C. Larvae removed from -80 ˚C were washed and dounced in FA Buffer (50 mM HEPES/KOH pH 7.5, 1 mM EDTA, 1% Triton X-100, 0.1 % sodium deoxycholate, 150 mM NaCl, filter sterilized) with 0.1% sarkosyl, 1 mM PMSF, and 1X protease inhibitors (protease inhibitor cocktail set I – Calbiochem, catalog # 539131). Sonication was performed in a Biopruptor ® Pico (catalog # B01060010) to obtain fragments between 200 and 800 bp in length (30 seconds on, 30 seconds off, 15 cycles) and checked by agarose gel electrophoresis. Protein concentration was determined using the Bradford assay (Bio-Rad, catalog # 500-0006). For chromatin immunoprecipitation, 1-2 mg of larval extract and 3-5 µg of antibody were used (see **supplemental file 1**), and allowed to rotate overnight at 4˚C. nProtein A Sepharose beads (Cytiva, catalog # 17528001) were added and incubation was allowed to continue for 2 hours. Beads were washed with FA buffer, FA buffer - 1 M NaCl, FA buffer - 500 mM NaCl, TEL buffer (0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0), and TE, respectively. Immunocomplexes were eluted with elution buffer (1% SDS, 250 mM NaCl, 10 mM Tris pH 8.0, 1 mM EDTA) at 65 ˚C for 15 min with mixing. Eluted immunocomplexes were treated with 2 µL proteinase K (10 mg/mL) for 1 hour at 50 ˚C, and then reverse crosslinked overnight at 65 ˚C. DNA was purified with the Qiagen MinElute PCR Purification Kit (catalog # 28006), and stored at -20 ˚C. Single-end (75 bp) sequencing was performed at the Genomics Core at the Center for Systems and Genome Biology, New York University using the Illumina NextSeq 500 platform.

ChIP-seq data processing and analysis

Reads were aligned with bowtie2 (version 2.4.2) using the default parameters. Sam files were converted to bam files, sorted and indexed using samtools (version 1.11). Bam files were used to generate bigwig files with the bamCoverage tool in deeptools (version 3.5.0) with the following parameters: --binsize 10 --minMappingQuality 20 --extendReads 200 --ignoreDuplicates --normalizeUsing CPM --exactScaling --ignoreForNormalization chrM. Enrichment was normalized to input with the bigwigCompare tool. Genome browser tracks were made with the hicPlotTADs tool in HiCExplorer (version 3.6). Enrichment around transcription start sites (TSS) and gene bodies were computed with the multiBigWigSummary tool in deeptools (version 3.5.0), and plotted in R with ggplot2 (version 3.5.0)[55].

RNA-seq analysis
We used the nf-core rna-seq pipeline (version 23.04.01) to generate counts and TPM tsv files. The workflow of the pipeline under our parameters was as follows: catenate technical replicates, infer strandedness with Salmon (version 1.10.1), asses the sequencing quality with FastQC (version 0.11.9), remove rRNA reads with sortMeRNA (version 4.3.4), align with STAR (version 2.7.10a), quantify with Salmon (version 1.10.1), and summarize counts and TPM quantification at the gene level with tximport (version 1.16.0). We used the follow flags: -profile singularity, --remove_ribo_rna, --save_non_ribo_reads, --save_reference, --skip_umi_extract, --skip_trimming, --skip_bbsplit_reads, --skip_biotype_qc, –skip_stringie, --skip_deseq2_qc. We ran DESeq2 (version 1.42.0) on the count tsv file in R, and used ggplot2 (version 3.5.0) for all plots[55,56]. We used the TPM tsv files to filter out unexpressed genes (mean TPM of replicates > 1 in both sexes).

Statistics
Wilcoxon rank sum tests were run in R using the wilcox.test function. Permutation tests were also run in R. In brief, autosomes were collapsed, and the observed difference in mean loop size between the autosomes (A) and the X chromosomes (X) was calculated as our test statistic. Mean loop size was defined as the separation in base pairs at the local maxima of the log derivative of the P(s)[57–59]. Permutations were run (n=10000) by extracting slope values, and randomly sampling a chromosome (A or X) without replacement. For each permutation, the mean loop size of A and X was selected, and the difference was calculated. A distribution of differences was plotted with ggplot2 (version 3.5.0). We calculated the p-value as a proportion of permuted test statistics that were more extreme than the observed test statistic.

Code availability
All scripts run in this study were deposited to the following Github repository: https://github.com/ercanlab/2024_Aharonoff_et_al.git.

Data availability
The data runs used and generated in this paper are provided in **supplemental file 1**. All data generated in this study were deposited to the Gene Expression Omnibus database under series numbers GSE267962 and GSE267963. The re-scaffolded *P. pacificus* genome and lifted annotation of protein coding genes can be found at https://github.com/ercanlab/2024_Aharonoff_et_al.git

RESULTS

Ortholog analysis of DCC subunits indicates divergence outside of *Caenorhabditis*

Nematodes represent a large number of species with distinct genomes. We used a recent phylogenomic analysis to choose a set of species from Rhabditina (clade V)[60] (**Figure 1B**). All species selected had BUSCO scores of at least 85% (**Table 1**). In all, we sampled from Pleiorhabditis, Anarhabditis, Synrhabditis, Dipoglastridae, and Strongylida, which display significant divergence in sex chromosome content (**Figure 1B**)[61–63]. Our outgroup was *Brugia*

malayi (Spirurina, clade III). To determine the conservation of the DCC, we searched for the orthologs with Orthofinder, which uses a reciprocal best blast hits (RBBH) approach.

The DCC subunits SDC-1, SDC-2, and SDC-3 function in both sex determination and dosage compensation[64]. It is well documented that sex determination mechanisms evolve with relative rapidity and are highly flexible to change[21,65]. SDC-2 is expressed specifically in XX embryos, and initiates hermaphrodite development and X chromosome dosage compensation[66]. SDC-3 physically interacts with SDC-2, helping repress the male determination gene *her-1* and recruit the DCC to the X chromosomes[64]. Consistent with their role as upstream regulators of sex determination, we only find orthologs of SDC-2 and SDC-3 in *Caenorhabditis* (**Figure 1B**).

Unlike SDC-2 and SDC-3, SDC-1 is not an essential gene, and is not required for DCC binding to the X chromosomes[67], but contributes to *her-1* repression[64,68–70]. We found SDC-1 orthologs in all species analyzed (**Figure 1B**). In *C. bovis, C. becei, C. elegans, C. briggsae, C. remanei,* and *M. japonica,* multiple orthologs were identified (**Supplemental file 1**). SDC-1 is a predicted DNA binding transcription factor with seven zinc fingers[71]. It is possible that the conservation and expansion of SDC-1 is related to its TF function in hermaphrodite differentiation.

The DPY-21 subunit of the DCC contains clear orthologs across all species analyzed, and DPY-30 in all but two (**Figure 1B**). DPY-21 is a conserved histone demethylase that is expressed in both the germline and soma with functions outside dosage compensation[72]. DPY-30 is a subunit of both the DCC and COMPASS, a chromatin modifying complex that is conserved across species[73–75]. Within COMPASS, DPY-30 interacts with ASH-2, and is required for trimethylation of H3K4, which is associated with active gene promoters[73,76,77]. The non-dosage compensation roles of DPY-21 and DPY-30 in chromatin regulation may contribute to their deeper conservation.

Phylogenetic analysis of SMC-4 reveals two independent duplications in *Caenorhabditis* and *Pristionchus*

In condensin I-DC, DPY-27 replaces SMC-4, an ATPase subunit of the canonical condensin I. To determine whether DPY-27 is a paralog of SMC-4, and to uncover the timing of the proposed duplication, we identified SMC-4 homologs and produced a maximum-likelihood tree of the sequences in the same sample set of species (**Figure 1C, Table 2**). Our results suggest that DPY-27 is an SMC-4 paralog and is at least as old as the common ancestor of *C. elegans* and *C. bovis*, but younger than the common ancestor shared between *Caenorhabditis* and *Oscheius* **(Figure 1C**).

Interestingly, we observed an independent duplication of SMC-4 in the *Pristionchus* lineage. To examine the paralogs, we performed a multiple sequence alignment of the N and C terminal ATPase domains, as well as regions that were previously found to be conserved and essential for SMC-4 ATPase function[78]. Similar to other SMC subunits, the ATPase activity of DPY-27 is essential for its function[79]. *C. elegans* DPY-27 has the conserved Walker A motif on the N terminus[28], along with site specific changes to the ancestral SMC-4 sequence (**Figure 1D**). In

Pristionchus pacificus and *exspectatus*, the Walker A motif is also conserved in both paralogs. Furthermore, we observe site specific changes independent of *Caenorhabditis* upstream of the Walker A motif (**Figure 1D**). We noticed that one group of paralogs in *P. pacificus* and *P. exspectatus* contains conserved changes to the ancestral SMC-4 sequence, suggesting that Iso_D.2636.1 and PREXSPEC000002716 diverged for a new function.

Hi-C analysis reveals X-specific topologically associating domains in *P. pacificus*

In *C. elegans*, condensin I-DC-mediated DNA loop extrusion increases 3D contacts, and forms loop-anchored TADs specifically on the X chromosomes[36,37]. A small number of cis-regulatory elements that function to recruit the DCC to the X chromosomes (recruitment elements on the X, *rex*) form the boundaries between TADs, and contact each other over hundreds of kilobase distances, forming *rex-rex* loops[36,37] (**Figure 2A**).

To independently validate the results of our phylogenetic analysis, we compared the Hi-C features of the X chromosomes to that of autosomes in five of our sampled species (**Figure 2A**). We obtained previously published Hi-C data in *C. elegans*[36], *C. remanei*[80] and *O. onirici* (Wellcome Sanger Institute), and generated data in *P. pacificus* and *O. tipulae*. In each, the X chromosomes are composed of a mosaic of ancestral segments[61–63]. Dosage compensation by condensin I-DC is found in the somatic cells of hermaphrodites and females. Therefore, we performed our Hi-C in early stage larvae, before germline cells begin to proliferate (see materials and methods).

While analyzing our Hi-C data mapped to the *P. pacificus* genome, we noticed several misjoins and inversions on chromosomes V and X (**Figure S1**). There were a total of 41 small, unplaced contigs (median length=43,467 bp), many of which displayed contacts with the X chromosome. With our HiC data, we were able to fix the misjoins and inversions on chromosomes V and X, and placed pbcontig517 on the left arm of the X chromosome (**Figure S1**). We also lifted over the annotations of protein coding genes. We re-mapped our *P. pacificus* Hi-C reads to the updated assembly.

In *C. elegans*, condensin I-DC forms loop-anchored TADs specifically on the hermaphrodite X chromosome **(Figure 2B, Figures S2-S3**). In *C. remanei*, all DCC subunits have clear orthologs, and TAD structures are observed specifically on the X chromosomes (**Figure 2B**). The less defined TAD features are expected because the *C. remanei* Hi-C data was obtained from mixed sex and mixed stage worms, diluting the hermaphrodite and soma-specific features of dosage compensation with male and germ cells.

Strikingly, *P. pacificus* also displayed loop anchored TADs on the X chromosome (**Figure 2B**). While *P. pacificus* has fewer TADs compared to *C. elegans* (**Figure S3**), its less assembled X chromosome may be affecting their detection. Autosomes did not show TADs in all species (**Figure 2B**). Therefore, in *C. elegans*, *C. remanei* and *P. Pacificus*, loop-anchored TADs are specific to the X chromosomes.

Distance decay analysis supports the presence of an X-specific loop extruding factor in *P. pacificus*

The loop extrusion activity of condensin I-DC on the X chromosomes is also detected by analyzing the decay of DNA contacts as a measure of increased distance, P(s)[36,81] (**Figure 3, top panel**). Previously, we showed that condensin I-DC mediated DNA loops produce a characteristic shoulder on the distance decay of 3D contacts from the X chromosome compared to the autosomes[36,57,81,82]. We performed the same analysis on our *P. pacificus* Hi-C data, and found that the X chromosome shows the characteristic shoulder not found on the autosomes (**Figure 3, Figure S4**).

Modeling of loop extrusion factors (LEFs) suggests that the local maxima of the log-derivative of the P(s) corresponds to the mean loop size, which is decreased by increasing the number of LEFs[57–59]. Indeed, the mean loop size of the *P. pacificus* X chromosome shows the same shift to the left observed in *C. elegans* (**Figure 3, Figure S4**). To test if the mean loop size was statistically different between the X chromosome and autosomes, we performed a permutation test, which supported the presence of an X-specific loop extruder in *P. pacificus* (**Figure S5**). As our test statistic, we collapsed the autosomes, and measured the difference in mean loop size between the autosomes and X chromosomes. In *P. pacificus* and *C. elegans*, the absolute value of the permuted test statistic was greater than the absolute value of the observed in 0.89% and 0.37% of permutations, respectively. In addition, a phylogenetic analysis of cohesin, condensin, and SMC-5/6 subunits found no other SMC duplications (**Figure S6**). Taken with the presence of loop-anchored TADs (**Figure 2B**) and a diverged SMC-4 variant (**Figure 1D**), our data suggest that *P. pacificus* contains a X chromosome specific condensin complex.

Hi-C analysis indicates no X-specific loop anchored TADs in *O. tipulae* and *O. onirici*

Unlike *P. pacificus*, *O. tipulae* does not contain paralogs of SMC-4 (**Figure 1C**). Importantly, our Hi-C data in *O. tipulae* did not display loop-anchored TADs on the X chromosome (**Figure 2B, Figures S2-S3**). Consistent with the lack of an X chromosome specific loop extruder, in *O. tipulae*, there was no difference between mean loop size on the X and autosomes (**Figure 3, Figure S4-S5**). We also analyzed the publicly available Hi-C dataset for *O. onirici* obtained from mixed stage, hermaphrodite animals (Wellcome Sanger Institute). While the mixed stage and mixed sex *C. remanei* data shows loop-anchored TADs, we observed no TADs in *O. onirici* (**Figure 2B, Figures S2-S3**). Together with our phylogenetic analysis, our Hi-C results place *Oscheius* in the lineage lacking an X specific loop extruder.

***O. tipulae* X chromosomes are dosage compensated**

Using mRNA-seq data from males and hermaphrodites, we previously established that *P. pacificus* and *C. remanei* X chromosomes are dosage compensated[83]. In *O. tipulae*, the absence of an SMC-4 paralog and loop-anchored TADs could either be because this species does not compensate or because it uses an alternate mechanism. To test if *O. tipulae* dosage

compensates, we analyzed publicly available RNA-seq data[84] and applied the same criteria established for assessing dosage compensation in *C. elegans*[83,85].

Dosage compensation mechanisms control the balance between the X and autosomes, and equalize X chromosomal transcript levels between sexes[18,86]. If there is dosage compensation, the average ratio of mRNA expression between sexes should be similar between the X chromosomes and autosomes. We analyzed mRNA-seq data in *B. malayi*, *O. tipulae*, and *H. contortus* adults, and re-analyzed data from *P. pacificus* young adults and *C. elegans* adults[83,84,87–89]. In each case, the average ratio of gene expression between sexes on the X and autosomes was similar (**Figure 4A, top panel**). The expected difference between X and autosomal expression in the absence of dosage compensation is observed in the larvae of a dosage compensation mutant *C. elegans* strain (*dpy-21(y428)*)[35] (**Figure 4B**).

In *C. elegans*, dosage compensation occurs in the somatic cells, but not in the germ cells, which can make up to two thirds of all cells in adults[85,90]. In addition, genes expressed in the germline are underrepresented on the X chromosomes due to non-sex-specific silencing during meiosis[91,92]. To evaluate dosage compensation in somatic cells, we used our previous strategy of enriching for soma expressed genes based on a cutoff selected by the bimodal distribution of differential expression between sexes[83].

In *C. elegans*, genes expressed specifically in the germ cells often show greater than 8-fold difference between hermaphrodites (or females) and males [83]. Thus, using a cutoff of log2 fold change=3 removes most germline-enriched genes[83]. Similar to *C. elegans*, the absolute log2 fold change between sexes shows a bimodal distribution in *O. tipulae* (**Figure 4C**). The level of male-biased gene enrichment on the X chromosome depends on the developmental timing of germline proliferation[83]. We observed varying degrees of male-biased gene depletion in *B. malayi* and *H. contortus*, which may be due to differences in genome evolution or germ cell proliferation (**Figure S7**). Regardless, using the same cutoff across species is a conservative approach to enrich for soma-expressed genes. This resulted in a log2 fold change between sexes to be closer to 0 (**Figure 4A, bottom panel**) and less variation between autosomes (**Figure S8**).

Each autosome harbors thousands of different genes with varied sex-biased gene expression (**Figure S7, S8**). We have used this variability to statistically test for the presence of X chromosome dosage compensation in *O. tipulae*. We employed a “one versus rest” approach, comparing the mean fold difference between sexes on each chromosome to the rest (e.g., I v.s. II, III, IV, V, and X). Similar to *C. elegans* and *P. pacificus*, in *O. tipulae* we found the difference in sex-biased gene expression between the X and autosomes was no more than the difference between an autosome and the rest of the chromosomes (**Figure 4D**). In contrast, in the *C. elegans* dosage compensation mutant (*dpy-21(e428)*), the difference between the X and autosomes is readily apparent. Therefore, X chromosomes are dosage compensated in *O. tipulae*. Together with *B. malayi* and *H. contortus*, these results suggest that dosage compensation is ancestral to condensin I-DC in nematodes, as expected by the model of sex chromosome evolution.

H4K20me1 is enriched on the X chromosomes in both *P. pacificus* and *O. tipulae*

Although our results indicate that there is dosage compensation in the analyzed nematode species, the mechanisms are not known. In *C. elegans*, X chromosome repression is mediated by condensin I-DC and DPY-21[34,72,93]. The demethylation of H4K20me2 by DPY-21 results in the enrichment of H4K20me1 on the X chromosomes compared to autosomes[34,72,93]. The repression of the X chromosomes also results in the depletion of histone modifications associated with active transcription[94].

We wondered if *P. pacificus* and *O. tipulae* show X-specific differences in histone marks associated with dosage compensation in *C. elegans*[34,93,94]. We performed ChIP-seq analysis of H4K20me1, H3K4me3, and as a negative control, IgG in J2 *P. pacificus* hermaphrodites, and early stage synchronized larvae in *O. tipulae*. We found that, similar to *C. elegans*, H4K20me1 is elevated across the entire X chromosome relative to autosomes in both species (**Figure 5A**).

H4K20me1 and H3K4me3 ChIP-seq data shown across a representative region highlights the gene-body and promoter enrichment of H4K20me1 and H3K4me3, respectively (**Figure 5B**). This suggests that the functions of these modifications are conserved. Analyzing relative enrichment of H4K20me1 on each chromosome indicated that H4K20me1 is enriched on the X chromosomal gene bodies (**Figure 5C, Figure S9**). We validated that the lower enrichment of H4K20me1 on the *O. tipulae* X chromosomes is statistically significant by comparing average gene-body ChIP-seq zscores in each chromosome to the rest (**Figure 5D**). The enrichment of H4K20me1 in *O. tipulae* is in contrast to the Hi-C data, where loop-anchored TADs are present in *P. pacificus*, but not *O. tipulae*. It is possible that *O. tipulae* has a mechanism of X chromosome repression that employs H4K20 monomethylation without condensin.

DISCUSSION

In this study, we found that *C. elegans* condensin I-DC evolved in the lineage leading to *Caenorhabditis* in the presence of an existing mechanism of dosage compensation (**Figure 6**). Said existing mechanism seems to have its basis in the conserved histone demethylase, DPY-21. Furthermore, a similar condensin based mechanism may have evolved in parallel in *Pristionchus*, which is indicative of common selective pressures acting on dosage compensation, and constraints on the evolution of dosage compensation in nematodes.

XO sex determination and dosage compensation are ancient in nematodes

The degree of Y chromosome degeneration in XY sex determination systems can be a powerful predictor of dosage compensation. For example, in the fish genus *Poecilia*, dosage compensation is only found in species with degenerative Y chromosomes[95]. In nematodes, the complete loss of the Y chromosome is ancestral, although the Y chromosome can re-evolve by fusion of the X chromosome to an autosome as observed in *B. malayi*[61]. Our analysis of mRNA-seq expression between sexes in *O. tipulae*, *H. contortus*, and *B. malayi* found that each of their X chromosomes are dosage compensated (**Figure 4**). A parsimonious view of the data

suggests that dosage compensation was present in the common ancestor of Rhabditida and Spirurina, and may be concurrently ancestral with the XO sex determination system (**Figure 6**).

Condensin-mediated dosage compensation evolved recently in the lineage leading to *Caenorhabditis*

Our phylogenetic analysis shows that condensin I-DC arose through duplication and divergence of SMC-4 in the lineage leading to *Caenorhabditis* (**Figure 1**). In *C. elegans*, the presence of looped anchored TADs are a molecular phenotype of condensin I-DC activity[36]. While X-specific TADs were found in *C. remanei*, no such TADs were found in *O. tipulae* and *O. onirici*, which lack an SMC-4 paralog (**Figure 2B**, **Figure 6**). Together, these results suggest that condensin I-DC is a newly evolved mechanism of dosage compensation.

Changes to sex chromosome composition is common in nematodes and may explain the inception of condensin I-DC in *Caenorhabditis*[61,63]. In *Drosophila*, autosome to sex chromosome fusions are followed by degeneration of the non-fused homolog, leaving the neo X-chromosomal genes in one copy[24,26,96]. This dosage problem was solved by evolution of DNA sequence motifs that recruit the existing dosage compensation complex to the neo-X chromosomes. In contrast, different mechanisms of dosage compensation seem to be acting on the neo and ancestral segments of the Z chromosome in the Monarch butterfly, *D. plexippus*[97]. In *D. plexippus*, while the neo-segment of the Z chromosome is upregulated in females (ZW) as in *D. melanogaster*, the ancestral segment of the Z chromosome is downregulated in males (ZZ) as in *C. elegans*[97]. Therefore, new mechanisms of X chromosome regulation may evolve in the presence of an old one.

Our data supports the idea that in *C. elegans*, condensin I-DC is a new mechanism of dosage compensation, evolving in the presence of an existing mechanism in the nematode clade. It is possible that an X chromosome event was the selective pressure to evolve condensin I-DC. However, such an event is not obvious from previous analyses of X chromosome evolution in nematodes[61–63]. Furthermore, while the X chromosomes are largely syntenic in *C. elegans* and *C. briggsae*, the DNA sequences that recruit the dosage compensation complex to the X chromosomes functionally diverged between the two species[27]. Therefore, large chromosomal events are not required for the continual evolution of the dosage compensation machinery.

Parallel evolution of condensin-mediated dosage compensation in *P. pacificus*

In perhaps the most surprising observation, we identified an independent duplication of SMC-4 in the common ancestor of *P. pacificus* and *P. exspectatus* (**Figure 1**, **Figure 6**). Hi-C analysis in *P. pacificus* revealed X-specific loop-anchored TADs in hermaphrodite larvae (**Figure 2B**). An exploration of the other SMC proteins, namely cohesin, condensin I, condensin II, and SMC-5/6 did not reveal any other duplication events in *Pristionchus* (**Figure S6**). A multiple sequence alignment of the conserved ATPase domains implicated one group of SMC-4 paralogs in dosage compensation (**Figure 1D**). The independent duplication of SMC-4 in *Caenorhabditis* and *Pristionchus* towards a possible dosage compensation function suggests a common

selective pressure in these two lineages. It will be important to determine what constraints lead to the parallel evolution of condensin-based mechanisms of dosage compensation in nematodes.

SMC genes are known to produce new genes by duplication. In fact, the entire SMC family of genes arose from a common SMC ancestor through duplication and divergence[98–101]. Interestingly, an SMC protein for dosage compensation has evolved in not one, but *two* clades with independent X chromosome ancestry. The first is the nematode condensin I-DC, and the second is SMCHD1 in mammals[102]. Similar to the *C. elegans* DPY-27, SMCHD1 functions in 3D organization and repression of the inactivated X chromosomes in mice and humans[102–106]. It is possible that the chromosome-wide binding capabilities of SMC complexes are well suited to solve the chromosome-wide transcriptional imbalance that necessitates dosage compensation.

Does an ancestral mechanism based on H4K20me1 predispose nematodes to evolve a condensin-based dosage compensation mechanism?

ChIP-seq data in *O. tipulae* shows that H4K20me1 is enriched on the X chromosome, despite the absence of an SMC-4 duplication and X-specific loop-anchored TADS. In *C. elegans*, H4K20me1 is enriched by DPY-21, which is recruited by condensin I-DC to the X chromosomes in mid embryogenesis[34,72,93]. H4K20me1 enhances X chromosome repression for dosage compensation, which is essential during late embryogenesis[35,72,107,108]. In *C. elegans*, DPY-21 also functions in the germline without condensin I-DC, where it is required for increasing H4K20me1on the autosomes, but not the X chromosome[72].

H4K20me1 is a highly conserved histone modification that is enriched on mitotic chromosomes when canonical condensins are active[109]. We speculate that the mechanistic link between H4K20me1 and condensin I contributed to the process of condensin DC evolution. Canonical condensin I remains cytoplasmic until the nuclear envelope breaks down during cell division[110,111]. Perhaps the nematode condensin I binds to the X chromosomes during interphase to increase H4K20me1, while functioning to compact all chromosomes during mitosis. Condensin I's dual role in dosage compensation and mitosis presents it with two possibly conflicting constraints. Duplication of the SMC-4 subunit followed by divergence would alleviate one of its constraints, allowing for the separate evolution of two condensin I variants.

Interestingly, an additional SMC may have evolved in *Caenorhabditis* to further mediate the potential interaction between condensin I and I-DC. Mass spectrometry analysis of proteins that interact with DPY-27 found a small SMC-like protein 1 (SMCL-1) in *C. elegans*[112]. SMCL-1 lacks the SMC hinge and coils, contains a non-functional ATPase domain, and negatively regulates condensin I and I-DC[112]. Future studies should address if parallel evolution of the SMC-4 paralog in *P. pacificus* also necessitated the emergence of a negative regulator of condensin I.

Concluding remarks

While the field has been able to assess whether X chromosomes are dosage compensated or not by using mRNA-seq analysis, the mechanisms of dosage compensation have been traditionally addressed in model organisms representing just a few clades. Recent efforts have been made to extend the characterization from mRNA-seq to chromatin modifications in less represented groups like butterflies, moths, mosquitoes, and lizards[97,113–116]. By combining a phylogenetic analysis of the *C. elegans* DCC with a comparative analysis of its epigenomic signatures, our work found that a new dosage compensation mechanism evolved in the presence of an ancestral one in nematodes. Surprisingly, this mechanism may have evolved independently in both *Caenorhabditis* and *Pristionchus*, which means the process of evolving dosage compensation may be constrained. The observation that SMC proteins repeatedly evolved for dosage compensation in nematodes and mammals argues for a fundamental link between the chromosome-wide activity of SMC complexes and the need to regulate transcription across the entire X chromosome. A more extensive comparison of dosage compensation mechanisms in nematodes will undoubtedly improve our understanding on the constraints of dosage compensation evolution and its draw for SMC protein evolution.

CONFLICTING INTERESTS

None.

AUTHOR CONTRIBUTIONS

A.A. and A.W. made the phylogenetic trees. A.A. and J.K. generated the HiC libraries. A.A. generated the ChIP libraries, analyzed the sequencing data, re-scaffolded and lifted over the *P. pacificus* genome, ran the statistical analyses, and prepared the figures. A.A. and S.E. conceived the project and wrote the manuscript.

ACKNOWLEDGEMENTS

We thank David Fitch and Karin Kiontke for their discussions on nematode phylogenetics. We thank Michael Werner (P), Bogdan Sierebriennikov (P), and Sophie Tintori (O) for their advice on working with *P. pacificus* and *O. tipulae*. We thank Christian Roedelsperger for discussions on *Pristionchus* phylogeny and genomes. The CEW1 strain was provided by the Caenorhabditis Genetics Center (CGC). Sequencing and processing of the raw data was done by the Genomics Core at the Center for Genomics and Systems Biology at New York University. AA was supported in part by NIGMS Predoctoral Fellowship T32HD007520, and AA and SE in part by R35GM130311. AW was supported by NSF REU Site award number 2148706.

Figure 1. Independent duplications of SMC-4 in *Caenorhabditis* and *Pristionchus* lead to the generation of novel SMC-4 proteins. (A) Condensin I-DC and five non-condensin proteins (green) make the *C. elegans* dosage compensation complex (DCC). Condensin DC shares all but one subunit, DPY-27, with the canonical Condensin I. (B) Not all orthologs of non-condensin proteins are found outside of *Caenorhabditis*. (C) A maximum-likelihood phylogeny of SMC-4 reveals that DPY-27 is an SMC-4 paralog that arose in the lineage leading to *Caenorhabditis* (orange circle). An independent duplication of SMC-4 occurred in *Pristionchus* (blue circle). Numbers indicate bootstrap values. (D) Multiple sequence alignment of the N-terminal ATPase domain shows conserved amino acid changes in DPY-27. Furthermore, independent amino acid changes separate the duplication events in *Caenorhabditis* and *Prisitionchus*.

Figure 2. X chromosome specific topologically associating domains are present in *P. pacificus*, but are absent in *O. tipulae and O. onirici*. (A) Schematic diagram of topologically associating domains (TADs) on the *C. elegans* X chromosome. TADs are defined by an increase in DNA-DNA contacts "within" relative to "between" segments of the genome. TADs are bordered by *rex* sites, reminiscent of mammalian CTCF sites, acting as both boundaries and loading sites of condensin I-DC, which mediate TADs by one-sided loop extrusion. (B) Hi-C matrices of representative X and autosome windows at 5 kb resolution with insulation scores. *P. pacificus* X chromosomes display TADs phenotypic of *C. elegans* dosage compensation. *O. tipulae* and *O. onirici* do not display X chromosome specific TADs. †, ††Data taken from Teterina et al. (2020) and the Wellcome Sanger Institute, respectively.

Figure 3. Distance decay curve in *P. pacificus* supports the presence of an X-specific loop extruder like the *C. elegans*-like dosage compensation condensin . Distance decay curves in each species show the contact probability, P(s), as a function of separation, s, for each chromosome at 5 kb resolution (top panel). Mean loop sizes for each chromosome were computed by taking the local maxima of the slope (bottom panel). Like in *C. elegans*, the *P. pacificus* hermaphrodite X chromosome has a smaller loop size than the autosomes, indicative of an X chromosome specific loop extruding factor. *O. tipulae* does not show a difference in mean loop size between the X chromosome and autosomes.

Figure 4. X chromosomes are consistently dosage compensated in multiple lineages of nematodes. (A) Log2 fold change between females/hermaphrodites and males in expressed genes on autosomes and the X chromosomes (top panel), and in soma enriched genes (bottom panel). (B) Log2 fold change between wild type *C. elegans* and a dosage compensation mutant, e428, showing the expected fold change on the X chromosome in the absence of dosage compensation. (C) Distribution of |log2(hermaphrodite/male)| in *O. tipulae*. Genes in *O. tipulae* display a bimodal distribution, separating the soma and germline enriched genes. (D) -log10(p-value) from Wilcoxon rank sum tests comparing the mean log2 fold change of chromosome to the mean of the rest (e.g., mean of X and mean of I-V). p-values in all but the e428 mutant cluster, suggesting no statistical difference between each chromosome in each wild type species. Inf = infinite.

Figure 5. *C. elegans*, *P. pacificus*, and *O. tipulae* X chromosomes are enriched for H4K20me1. (A) H4K20me1 ChIP-seq tracks (ChIP - input) of the X chromosome and a representative autosome in *P. pacificus*, *C. elegans*, and *O. tipulae*. (B) H4K20me1, H3K4me3, and IgG ChIP-seq tracks (ChIP - input) of 50-100 kb regions on the X chromosome. (C) z-score of log2(ChIP/input) H4K20me1 enrichment in gene bodies separated by chromosome. H4K20me1 is enriched in all three species, suggestive of a common ancestral strategy of dosage compensation in nematodes.

Figure 6. Model for the evolution of dosage compensation in nematodes. Dosage compensation is as old as the XO sex determination in nematodes, which is supported by its presence in *Brugia*, *Pristionchus*, *Caenorhabditis*, *Haemonchus*, and *Oscheius*. DPY-27, the condensin I-DC specific subunit, is an SMC-4 paralog found in *Caenorhabditis*. The duplication event occurred in the lineage leading to *Caenorhabditis* (orange circle). A second, independent SMC-4 duplication occurred in *Pristionchus* (dark blue circle). X chromosome specific TADs in hermaphrodites are phenotypic of condensin-mediated dosage compensation, and are found in *C. elegans*, *C. remanei*, and *P. pacificus*, but not in *O. tipulae* and *O. onirici*. Condensin-based systems may have evolved parallelly in *Pristionchus*. Enrichment of the repressive histone mark, H4K20me1 on the hermaphrodite X chromosome, is found even in non-condensin based dosage compensation systems, which is indicative of an ancestral mechanism common to Rhabditina (light blue circle). -, not found; *blank space*, not checked; , dosage compensation; , X specific TADS; no TADs on X chromosome; , X specific enrichment of H4K20me1.

Figure S1. Fixed misjoins and inversions on *P. pacificus* genome. Misjoins and inversions (left panel) on chromosome V and X of the *P. pacificus* genome were fixed (right panel) with Juicebox Assembly Tools, and pbcontig517 was placed on the left arm of chromosome X (arrow) using the scaffolding tool YaHS.

Figure S2. Whole chromosome Hi-C matrices. Hi-C matrices of chromosomes IV and X at 20-25 kb resolution for *P. pacificus*, *C. elegans*, *C. remanei*, *O. tipulae*, and *O. onirici*. X chromosomes do not display arms and centers, but in *C. elegans*, *C. remanei*, and *P. pacificus*, TADs are observed.

Figure S3. Replicates of Hi-C matrices and insulation scores. Replicates of Hi-C matrices and insulation scores for data generated in this study exhibit reproducibility in displaying X chromosome specific TADs.

Figure S4. Replicates of distance decay curves. Replicates of P(s) curve and its derivative exhibit reproducibility of estimated mean loop size.

Figure S5. The observed difference in mean loop size between the autosomes and X chromosome are significantly different in *P. pacificus* and *C. elegans*. Permutations were performed by randomly permuting the slope values between the X chromosome and autosomes, extracting the corresponding separation in bp, and subtracting the X from the A as

our test statistic. The distribution represents n=10000 permutations. Red line, observed difference; p, proportion of permutations that were more extreme than the observed difference.

Figure S6. Maximum-likelihood trees of SMC proteins in nematodes. Maximum likelihood trees of all condensin, cohesin, and SMC-5/6 subunits show no duplications in *P. pacificus*. Our trees were built with the all of the species used for the DCC specific subunits except for *A. rhodensis*, *N. americanus*, and *A. ceylanicum*

Figure S7. Distribution of |log2 fold change| in *B. malayi*, *P. pacificus*, *C. elegans*, and *H. contortus*. The distribution of differentially expressed genes between sexes is split into soma (log2 < 3) and sex-biased, germline enriched genes (log2 > 3). The dearth of germline enriched genes in *P. pacificus* was addressed in Albritton et al. (2016)[83].

Figure S8. log2 fold change plots separated by chromosome. Log2 fold change between female/hermaphrodite and males of soma enriched expressed genes in *B. malayi*, *P. pacificus*, *C. elegans*, *O. tipulae*, and *H. contortus* show little to no difference between chromosomes. PAR, pseudoautosomal region.

Figure S9. Replicates of ChIP-seq enrichment of H4K20me1 at gene bodies. Replicates of ChIP-seq z-score analysis exhibit reproducibility. IgG is shown as a negative control.

REFERENCES

1. Charlesworth, B. & Charlesworth, D. A Model for the Evolution of Dioecy and Gynodioecy. *The American Naturalist* **112**, 975–997 (1978).
2. Hodgkin, J. Genetic sex determination mechanisms and evolution. *Bioessays* **14**, 253–261 (1992).
3. Charlesworth, B. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* **6**, 149–162 (1996).
4. Jablonka, E. & Lamb, M. J. The evolution of heteromorphic sex chromosomes. *Biol Rev Camb Philos Soc* **65**, 249–276 (1990).
5. Charlesworth, B. The evolution of sex chromosomes. *Science* **251**, 1030–1033 (1991).
6. Rice, W. R. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116**, 161–167 (1987).
7. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**, 113–124 (2013).
8. Oliver, B. Sex, Dose, and Equality. *PLoS Biol* **5**, e340 (2007).
9. Sahakyan, A., Yang, Y. & Plath, K. The Role of Xist in X-Chromosome Dosage Compensation. *Trends Cell Biol* **28**, 999–1013 (2018).
10. Loda, A. & Heard, E. Xist RNA in action: Past, present, and future. *PLoS Genet* **15**, e1008333 (2019).
11. Brockdorff, N., Bowness, J. S. & Wei, G. Progress toward understanding chromosome silencing by Xist RNA. *Genes Dev* **34**, 733–744 (2020).
12. Straub, T., Gilfillan, G. D., Maier, V. K. & Becker, P. B. The Drosophila MSL complex activates the transcription of target genes. *Genes Dev* **19**, 2284–2288 (2005).
13. Lucchesi, J. C. & Kuroda, M. I. Dosage compensation in Drosophila. *Cold Spring Harb Perspect Biol* **7**, a019398 (2015).

14. Meyer, B. J. & Casson, L. P. Caenorhabditis elegans compensates for the difference in X chromosome dosage between the sexes by regulating transcript levels. *Cell* **47**, 871–881 (1986).

15. Albritton, S. E. & Ercan, S. Caenorhabditis elegans Dosage Compensation: Insights into Condensin-Mediated Gene Regulation. *Trends Genet* **34**, 41–53 (2018).

16. Chuang, P. T., Albertson, D. G. & Meyer, B. J. DPY-27:a chromosome condensation protein homolog that regulates C. elegans dosage compensation through association with the X chromosome. *Cell* **79**, 459–474 (1994).

17. Csankovszki, G. *et al.* Three distinct condensin complexes control C. elegans chromosome dynamics. *Curr Biol* **19**, 9–19 (2009).

18. Ercan, S. Mechanisms of x chromosome dosage compensation. *J Genomics* **3**, 1–19 (2015).

19. Landeen, E. L. & Presgraves, D. C. Evolution: From Autosomes to Sex Chromosomes — and Back. *Current Biology* **23**, R848–R850 (2013).

20. Furman, B. L. S. *et al.* Sex Chromosome Evolution: So Many Exceptions to the Rules. *Genome Biol Evol* **12**, 750–763 (2020).

21. Bachtrog, D. *et al.* Sex determination: why so many ways of doing it? *PLoS Biol* **12**, e1001899 (2014).

22. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol* **13**, e1002078 (2015).

23. Myosho, T., Takehana, Y., Hamaguchi, S. & Sakaizumi, M. Turnover of Sex Chromosomes in Celebensis Group Medaka Fishes. *G3 (Bethesda)* **5**, 2685–2691 (2015).

24. Ellison, C. & Bachtrog, D. Contingency in the convergent evolution of a regulatory network: Dosage compensation in Drosophila. *PLoS Biol* **17**, e3000094 (2019).

25. Ellison, C. E. & Bachtrog, D. Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* **342**, 846–850 (2013).

26. Bone, J. R. & Kuroda, M. I. Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila. *Genetics* **144**, 705–713 (1996).

27. Yang, Q. *et al.* X-chromosome target specificity diverged between dosage compensation mechanisms of two closely related Caenorhabditis species. *Elife* **12**, e85413 (2023).

28. Harvey, S. H., Krien, M. J. E. & O'Connell, M. J. Structural maintenance of chromosomes (SMC) proteins, a family of conserved ATPases. *Genome Biol* **3**, REVIEWS3003 (2002).

29. Hirano, T. Condensins: organizing and segregating the genome. *Curr Biol* **15**, R265-275 (2005).

30. Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R. & Meyer, B. J. C. elegans condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev* **16**, 729–742 (2002).

31. Strunnikov, A. V. & Jessberger, R. Structural maintenance of chromosomes (SMC) proteins: conserved molecular properties for multiple biological functions. *Eur J Biochem* **263**, 6–13 (1999).

32. Hirano, T. Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev* **26**, 1659–1678 (2012).

33. Csankovszki, G. Condensin function in dosage compensation. *Epigenetics* **4**, 212–215 (2009).

34. Vielle, A. *et al.* H4K20me1 contributes to downregulation of X-linked genes for C. elegans dosage compensation. *PLoS Genet* **8**, e1002933 (2012).

35. Kramer, M. *et al.* Developmental Dynamics of X-Chromosome Dosage Compensation by the DCC and H4K20me1 in C. elegans. *PLOS Genetics* **11**, 1005698 (2015).

36. Kim, J. *et al.* Condensin DC loads and spreads from recruitment sites to create loop-anchored TADs in C. elegans. *Elife* **11**, e68745 (2022).

37. Anderson, E. C. *et al.* X Chromosome Domain Architecture Regulates Caenorhabditis elegans Lifespan but Not Dosage Compensation. *Dev Cell* **51**, 192-207 6 (2019).

38. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Molecular Biology and Evolution* **35**, 543–548 (2018).

39. Manni, M., Berkeley, M. R., Seppey, M. & Zdobnov, E. M. BUSCO: Assessing Genomic Data Quality and Beyond. *Current Protocols* **1**, e323 (2021).

40. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).

41. Rödelsperger, C. The community-curated Pristionchus pacificus genome facilitates automated gene annotation improvement in related nematodes. *BMC Genomics* **22**, 216 (2021).

42. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157 (2015).

43. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).

44. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **30**, 772–780 (2013).

45. Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* **32**, 268–274 (2015).

46. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).

47. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:

Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).

48. Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: Flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol Evol* **14**, e10873 (2024).

49. Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).

50. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).

51. Shumate, A. & Salzberg, S. L. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

52. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

53. Open2C *et al.* Cooltools: enabling high-resolution Hi-C analysis in Python. Preprint at https://doi.org/10.1101/2022.10.31.514564 (2022).

54. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).

55. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

57. Gassler, J. *et al.* A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J* **36**, 3600–3618 (2017).

58. Polovnikov, K. E. *et al.* Crumpled Polymer with Loops Recapitulates Key Features of Chromosome Organization. *Phys. Rev. X* **13**, 041029 (2023).

59. Goloborodko, A., Marko, J. F. & Mirny, L. A. Chromosome Compaction by Active Loop Extrusion. *Biophysical Journal* **110**, 2162–2168 (2016).

60. Ahmed, M. *et al.* Phylogenomic Analysis of the Phylum Nematoda: Conflicts and

Congruences With Morphology, 18S rRNA, and Mitogenomes. *Front. Ecol. Evol.* **9**, 769565 (2022).

61. Foster, J. M. *et al.* Sex chromosome evolution in parasitic nematodes of humans. *Nat Commun* **11**, 1964 (2020).

62. Tandonnet, S. *et al.* Chromosome-Wide Evolution and Sex Determination in the Three-Sexed Nematode Auanema rhodensis. *G3 (Bethesda)* **9**, 1211–1230 (2019).

63. Gonzalez de la Rosa, P. M. *et al.* A telomere-to-telomere assembly of Oscheius tipulae and the evolution of rhabditid nematode chromosomes. *G3 (Bethesda)* **11**, jkaa020 (2021).

64. Chu, D. S. *et al.* A molecular link between gene-specific and chromosome-wide transcriptional repression. *Genes Dev* **16**, 796–805 (2002).

65. Hodgkin, J. Exploring the envelope. Systematic alteration in the sex-determination system of the nematode caenorhabditis elegans. *Genetics* **162**, 767–780 (2002).

66. Dawes, H. E. *et al.* Dosage compensation proteins targeted to X chromosomes by a determinant of hermaphrodite fate. *Science* **284**, 1800–1804 (1999).

67. Chuang, P. T., Lieb, J. D. & Meyer, B. J. Sex-specific assembly of a dosage compensation complex on the nematode X chromosome. *Science* **274**, 1736–1739 (1996).

68. Villeneuve, A. M. & Meyer, B. J. sdc-1: a link between sex determination and dosage compensation in C. elegans. *Cell* **48**, 25–37 (1987).

69. Villeneuve, A. M. & Meyer, B. J. The role of sdc-1 in the sex determination and dosage compensation decisions in Caenorhabditis elegans. *Genetics* **124**, 91–114 (1990).

70. Meyer, B. J. Mechanisms of sex determination and X-chromosome dosage compensation. *Genetics* **220**, iyab197 (2022).

71. Nonet, M. L. & Meyer, B. J. Early aspects of Caenorhabditis elegans sex determination and dosage compensation are regulated by a zinc-finger protein. *Nature* **351**, 65–68 (1991).

72. Brejc, K. *et al.* Dynamic Control of X Chromosome Conformation and Repression by a Histone H4K20 Demethylase. *Cell* **171**, 85-102.e23 (2017).

73. Pferdehirt, R. R., Kruesi, W. S. & Meyer, B. J. An MLL/COMPASS subunit functions in the C. elegans dosage compensation complex to target X chromosomes for transcriptional regulation of gene expression. *Genes Dev* **25**, 499–515 (2011).

74. Ernst, P. & Vakoc, C. R. WRAD: enabler of the SET1-family of H3K4 methyltransferases. *Brief Funct Genomics* **11**, 217–226 (2012).

75. Morgan, M. A. J. & Shilatifard, A. Reevaluating the roles of histone-modifying enzymes and their associated chromatin modifications in transcriptional regulation. *Nat Genet* **52**, 1271–1281 (2020).

76. Beurton, F. *et al.* Physical and functional interaction between SET1/COMPASS complex component CFP-1 and a Sin3S HDAC complex in C. elegans. *Nucleic Acids Res* **47**, 11164–11180 (2019).

77. Simonet, T., Dulermo, R., Schott, S. & Palladino, F. Antagonistic functions of SET-2/SET1 and HPL/HP1 proteins in C. elegans development. *Dev Biol* **312**, 367–383 (2007).

78. Hirano, M. & Hirano, T. Opening closed arms: long-distance activation of SMC ATPase by hinge-DNA interactions. *Mol Cell* **21**, 175–186 (2006).

79. Breimann, L. *et al.* The histone H4 lysine 20 demethylase DPY-21 regulates the dynamics of condensin DC binding. *J Cell Sci* **135**, jcs258818 (2022).

80. Teterina, A. A., Willis, J. H. & Phillips, P. C. Chromosome-Level Assembly of the Caenorhabditis remanei Genome Reveals Conserved Patterns of Nematode Genome Organization. *Genetics* **214**, 769–780 (2020).

81. Morao, A. K., Kim, J., Obaji, D., Sun, S. & Ercan, S. Topoisomerases I and II facilitate condensin DC translocation to organize and repress X chromosomes in C. elegans. *Mol Cell* **82**, 4202-4217.e5 (2022).

82. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).

83. Albritton, S. E. *et al.* Sex-biased gene expression and evolution of the x chromosome in nematodes. *Genetics* **197**, 865–883 (2014).

84. Dockendorff, T. C. *et al.* The nematode Oscheius tipulae as a genetic model for programmed DNA elimination. *Curr Biol* **32**, 5083-5098.e6 (2022).

85. Kramer, M., Rao, P. & Ercan, S. Untangling the Contributions of Sex-Specific Gene Regulation and X-Chromosome Dosage to Sex-Biased Gene Expression in Caenorhabditis elegans. *Genetics* **204**, 355–369 (2016).

86. Deng, X. *et al.* Evidence for compensatory upregulation of expressed X-linked genes in mammals, Caenorhabditis elegans and Drosophila melanogaster. *Nat Genet* **43**, 1179–1185 (2011).

87. Chung, M. *et al.* Multispecies Transcriptomics Data Set of Brugia malayi, Its Wolbachia Endosymbiont wBm, and Aedes aegypti across the B. malayi Life Cycle. *Microbiol Resour Announc* **7**, e01306-18 (2018).

88. Doyle, S. R. *et al.* Genomic and transcriptomic variation defines the chromosome-scale assembly of Haemonchus contortus, a model gastrointestinal worm. *Commun Biol* **3**, 656 (2020).

89. Laing, R. *et al.* The genome and transcriptome of Haemonchus contortus, a key model parasite for drug and vaccine discovery. *Genome Biol* **14**, R88 (2013).

90. Strome, S., Kelly, W. G., Ercan, S. & Lieb, J. D. Regulation of the X Chromosomes in Caenorhabditis elegans. *Cold Spring Harb Perspect Biol* **6**, 018366 (2014).

91. Reinke, V., Gil, I. S., Ward, S. & Kazmer, K. Genome-wide germline-enriched and sex-biased expression profiles in Caenorhabditis elegans. *Development* **131**, 311–323 (2004).

92. Wang, X. *et al.* Identification of genes expressed in the hermaphrodite germ line of C. elegans using SAGE. *BMC Genomics* **10**, 213 (2009).

93. Wells, M. B., Snyder, M. J., Custer, L. M. & Csankovszki, G. Caenorhabditis elegans

dosage compensation regulates histone H4 chromatin state on X chromosomes. *Mol Cell Biol* **32**, 1710–1719 (2012).

94. Street, L. A. *et al.* Binding of an X-Specific Condensin Correlates with a Reduction in Active Histone Modifications at Gene Regulatory Elements. *Genetics* **212**, 729–742 (2019).

95. Metzger, D. C. H., Sandkam, B. A., Darolti, I. & Mank, J. E. Rapid Evolution of Complete Dosage Compensation in Poecilia. *Genome Biol Evol* **13**, evab155 (2021).

96. Marín, I., Franke, A., Bashaw, G. J. & Baker, B. S. The dosage compensation system of Drosophila is co-opted by newly evolved X chromosomes. *Nature* **383**, 160–163 (1996).

97. Gu, L. *et al.* Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. *Curr Biol* **29**, 4071-4077.e3 (2019).

98. Van Hooff, J. J. E., Raas, M. W. D., Tromer, E. C. & Eme, L. Shaping up genomes: Prokaryotic roots and eukaryotic diversification of SMC complexes. Preprint at https://doi.org/10.1101/2024.01.07.573240 (2024).

99. Hirano, T. The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesion, and repair. *Genes Dev* **16**, 399–414 (2002).

100. Yoshinaga, M. & Inagaki, Y. Ubiquity and Origins of Structural Maintenance of Chromosomes (SMC) Proteins in Eukaryotes. *Genome Biol Evol* **13**, evab256 (2021).

101. Cobbe, N. & Heck, M. M. S. The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol* **21**, 332–347 (2004).

102. Blewitt, M. E. *et al.* SmcHD1, containing a structural-maintenance-of-chromosomes hinge domain, has a critical role in X inactivation. *Nat Genet* **40**, 663–669 (2008).

103. Sakakibara, Y. *et al.* Role of SmcHD1 in establishment of epigenetic states required for the maintenance of the X-inactivated state in mice. *Development* **145**, dev166462 (2018).

104. Bowness, J. S. *et al.* Xist-mediated silencing requires additive functions of SPEN and Polycomb together with differentiation-dependent recruitment of SmcHD1. *Cell Rep* **39**, 110830 (2022).

105. Wang, C.-Y., Colognori, D., Sunwoo, H., Wang, D. & Lee, J. T. PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. *Nat Commun* **10**, 2950 (2019).

106. Gdula, M. R. *et al.* The non-canonical SMC protein SmcHD1 antagonises TAD formation and compartmentalisation on the inactive X chromosome. *Nat Commun* **10**, 30 (2019).

107. Davis, T. L. & Meyer, B. J. SDC-3 coordinates the assembly of a dosage compensation complex on the nematode X chromosome. *Development* **124**, 1019–1031 (1997).

108. Plenefisch, J. D., DeLong, L. & Meyer, B. J. Genes that implement the hermaphrodite mode of dosage compensation in Caenorhabditis elegans. *Genetics* **121**, 57–76 (1989).

109. Beck, D. B., Oda, H., Shen, S. S. & Reinberg, D. PR-Set7 and H4K20me1: at the crossroads of genome integrity, cell cycle, chromosome condensation, and transcription. *Genes Dev* **26**, 325–337 (2012).

110. Shintomi, K. & Hirano, T. The relative ratio of condensin I to II determines chromosome shapes. *Genes Dev* **25**, 1464–1469 (2011).

111. Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of Condensins I and II in mitotic chromosome assembly in human cells. *Mol Biol Cell* **15**, 3296–3308 (2004).

112. Chao, L. F.-I., Singh, M., Thompson, J., Yates, J. R. & Hagstrom, K. A. An SMC-like protein binds and regulates Caenorhabditis elegans condensins. *PLoS Genet* **13**, e1006614 (2017).

113. Krzywinska, E. *et al.* femaleless Controls Sex Determination and Dosage Compensation Pathways in Females of Anopheles Mosquitoes. *Curr Biol* **31**, 1084-1091.e4 (2021).

114. Kalita, A. I. *et al.* The sex-specific factor SOA controls dosage compensation in Anopheles mosquitoes. *Nature* **623**, 175–182 (2023).

115. Marin, R. *et al.* Convergent origination of a Drosophila-like dosage compensation mechanism in a reptile lineage. *Genome Res* **27**, 1974–1987 (2017).

116. Rosin, L. F., Chen, D., Chen, Y. & Lei, E. P. Dosage compensation in Bombyx mori is achieved by partial repression of both Z chromosomes in males. *Proc Natl Acad Sci U S A* **119**, e2113374119 (2022).

TABLE 1

Table 1. Species sampled for phylogenetic analysis of dosage compensation proteins

Species	Family	Clade	BUSCO (S/D/F/M)
Haemonchus contortus	Strongylida	-	85/11/1/3
Ancylostoma ceylanicum	Strongylida	-	86/9/2/3
Necator americanus	Strongylida	-	46/51/1/2
Oscheius onirici	Rhabditidae	Synrhabditis	89/3/2/6
Oscheius tipulae	Rhabditidae	Synrhabditis	88/3/2/8
Auanema rhodensis	Rhabditidae	Synrhabditis	77/13/1/9
Caenorhabditis remanei	Rhabditidae	Anarhabditis	95/4/1/0
Caenorhabditis briggsae	Rhabditidae	Anarhabditis	80/19/0/0
Caenorhabditis elegans	Rhabditidae	Anarhabditis	66/34/0/0
Caenorhabditis becie	Rhabditidae	Anarhabditis	89/9/1/2
Caenorhabditis bovis	Rhabditidae	Anarhabditis	77/17/1/6
Mesorhabditis belari	Rhabditidae	Pleiorhabditis	72/13/2/13
Pristionchus exspectatus	Diplogastridae	-	90/3/4/4
Pristionchus pacificus	Diplogastridae	-	96/2/2/0
Pristionchus fissidentatus	Diplogastridae	-	87/2/7/4
Micoletzkya japonica	Diplogastridae	-	88/2/5/5
Brugia malayi	Onchocercidae	-	70/31/0/1

S, Complete and single copy. *D,* Complete and duplicated. *F,* Fragmented. *M,* Missing. Completeness is measured by summing *S* and *D*.

TABLE 2

Table 2. Homologs of *C. elegans* SMC-4

Species	Homolog
Haemonchus contortus	HCON_00069750
Ancylostoma ceylanicum	Acey_s0467.v2.g16259
Necator americanus	Necator_chrIII.pre1.g12501
Oscheius onirici	g14665
Oscheius tipulae	Oti_478_g
Auanema rhodensis	JU1783_LOCUS7082
Caenorhabditis remanei	CRE20351 CRE24237 CRE20350* CRE15073*
Caenorhabditis briggsae	CBG18034 CBG20143
Caenorhabditis elegans	F35G12.8 R13G10.1
Caenorhabditis becei	CSP29.g14748 CSP29.g16291
Caenorhabditis bovis	CBOVI.g4439 CBOVI.g4612
Mesorhabditis belari	mbelari.g20448
Pristionchus exspectatus	PREXSPEC000002716 PREXSPEC000004283 PREXSPEC000022998*
Pristionchus pacificus	ppa_stranded_DN27290_c2_g1_i1 Iso_D.2636.1
Pristionchus fissidentatus	PRIFISSI000001848
Micoletzkya japonica	MIJAPON000004369**
Brugia malayi	WBGene00234061

DPY-27 orthologs are colored orange. SMC-4 paralogs in *Pristionchus* are colored blue.
*These orthologs did not meet the sequence cover threshold of 50%, and were not used in downstream analyses.
**MIJAPON000005949 was also called as an ortholog. Upon inspection, both are segments of the same gene and their separation is a result of misannotation. Only MIJAPON000004369 was used in downstream analyes.

FIGURE 1

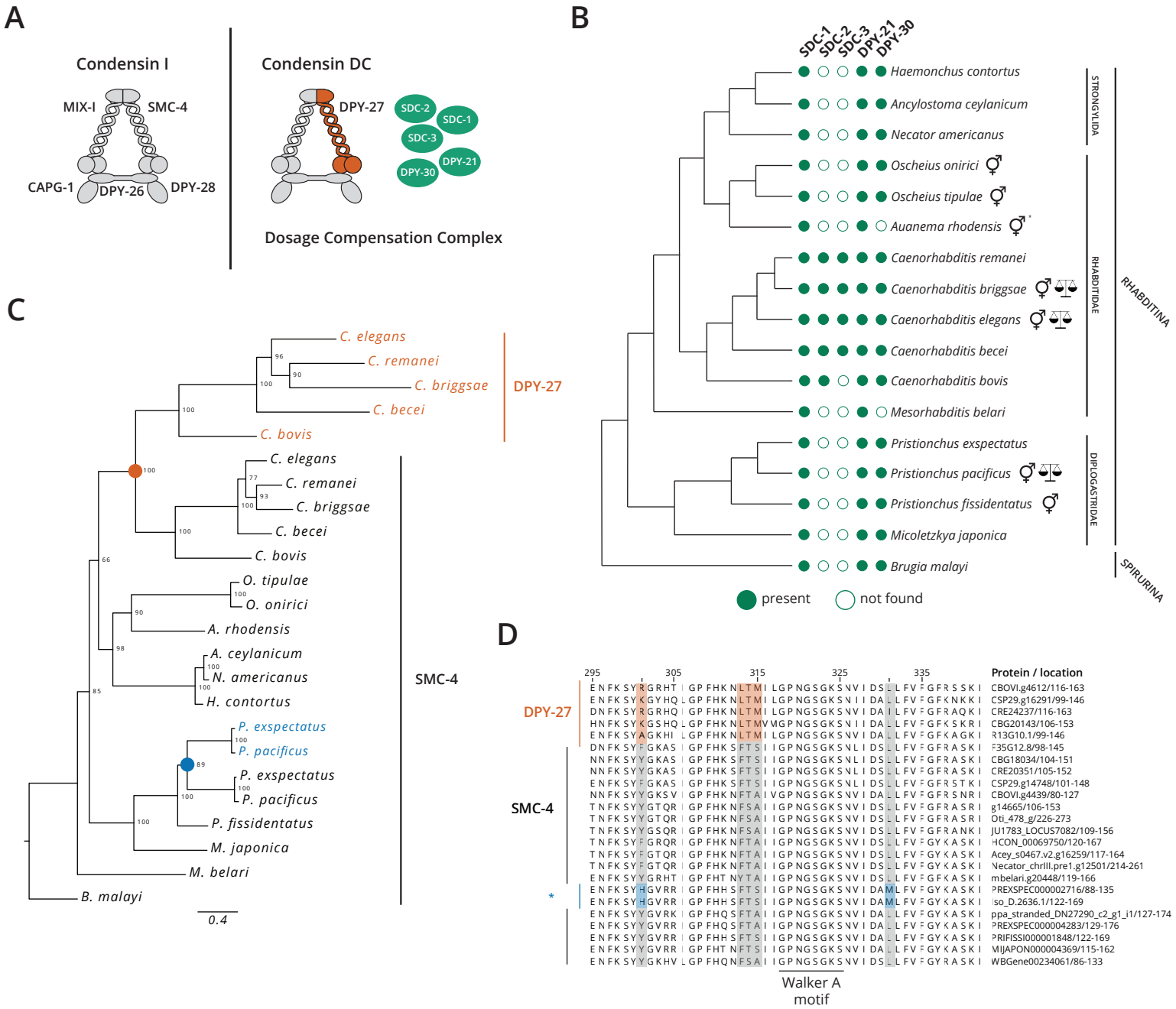

A
Condensin I
MIX-I
SMC-4
CAPG-1
DPY-26
DPY-28
Condensin DC
DPY-27
SDC-2
SDC-1
SDC-3
DPY-21
DPY-30
Dosage Compensation Complex
B
SDC-1
SDC-2
SDC-3
DPY-21
DPY-30
Haemonchus contortus
Ancylostoma ceylanicum
Necator americanus
Oscheius onirici
Oscheius tipulae
Auanema rhodensis
Caenorhabditis remanei
Caenorhabditis briggsae
Caenorhabditis elegans
Caenorhabditis becei
Caenorhabditis bovis
Mesorhabditis belari
Pristionchus exspectatus
Pristionchus pacificus
Pristionchus fissidentatus
Micoletzkya japonica
Brugia malayi
STRONGYLIDA
RHABDITIDAE
DIPLOGASTRIDAE
RHABDITINA
SPIRURINA
present
not found
C
C. elegans
C. remanei
C. briggsae
C. becei
C. bovis
DPY-27
O. tipulae
O. onirici
A. rhodensis
A. ceylanicum
N. americanus
H. contortus
P. exspectatus
P. pacificus
P. fissidentatus
M. japonica
M. belari
B. malayi
SMC-4
0.4
D
Protein / location
DPY-27
SMC-4
Walker A motif

FIGURE 2

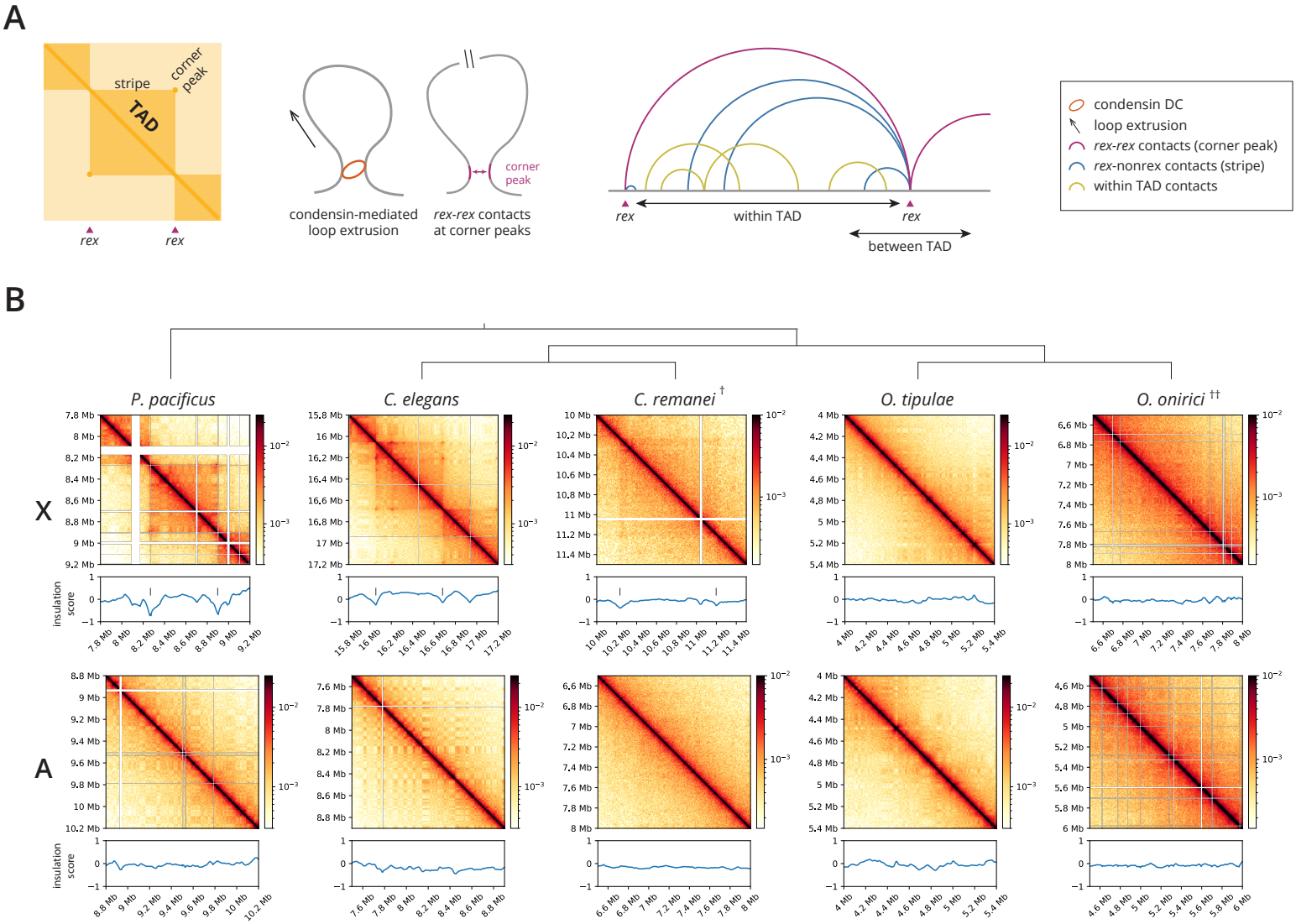

A
stripe
corner peak
TAD
rex
rex
condensin-mediated loop extrusion
rex-rex contacts at corner peaks
corner peak
rex
within TAD
rex
between TAD
condensin DC
loop extrusion
rex-rex contacts (corner peak)
rex-nonrex contacts (stripe)
within TAD contacts
B
P. pacificus
C. elegans
C. remanei †
O. tipulae
O. onirici ††
X
A
insulation score

FIGURE 3

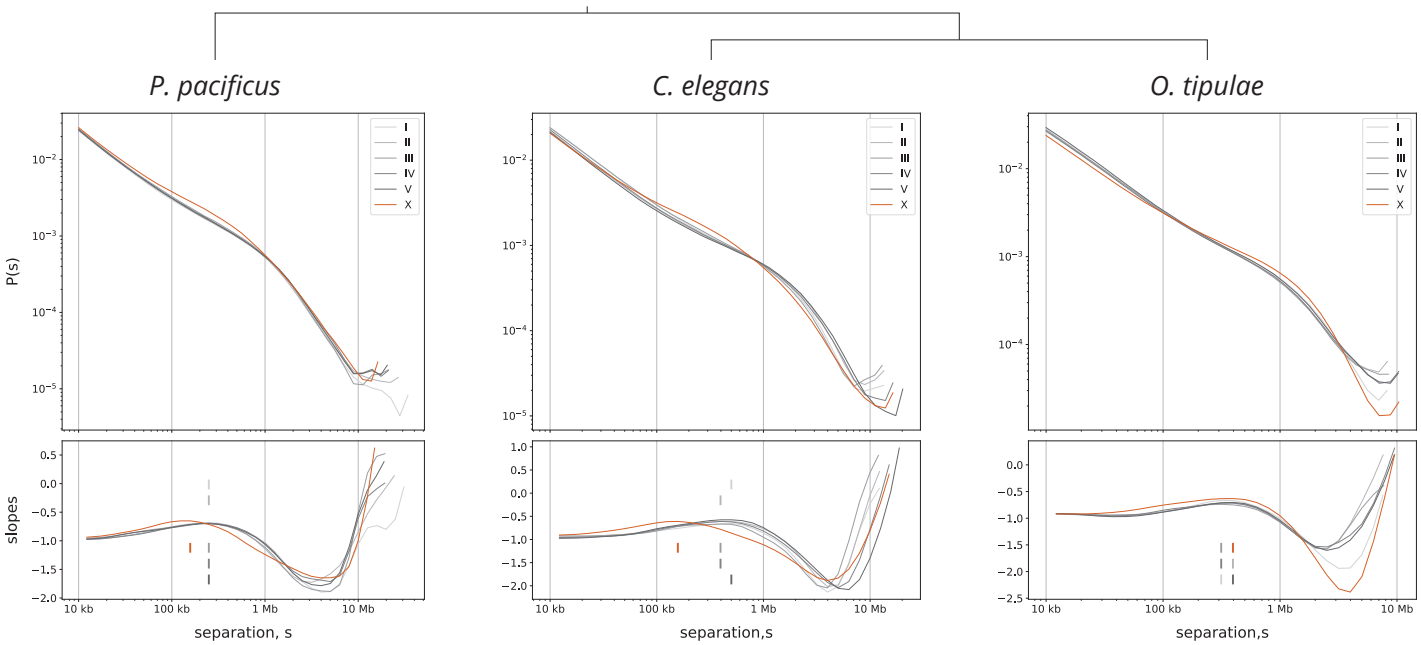
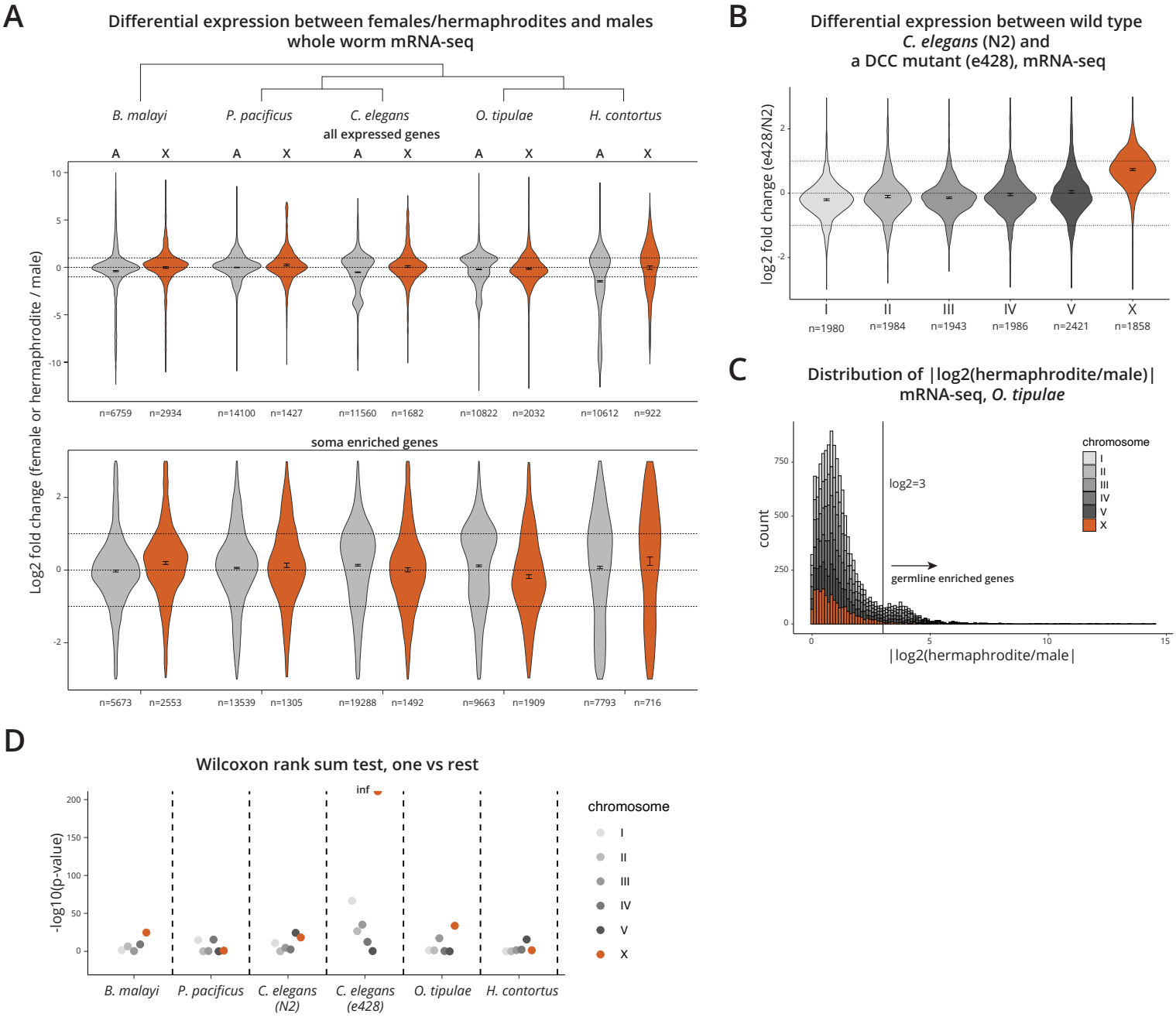

P. pacificus
C. elegans
O. tipulae
P(s)
slopes
I
II
III
IV
V
X
10 kb
100 kb
1 Mb
10 Mb
separation, s
separation,s

FIGURE 4


A
Differential expression between females/hermaphrodites and males
whole worm mRNA-seq
B. malayi
P. pacificus
C. elegans
O. tipulae
H. contortus
all expressed genes
A X A X A X A X A X
Log2 fold change (female or hermaphrodite / male)
n=6759 n=2934 n=14100 n=1427 n=11560 n=1682 n=10822 n=2032 n=10612 n=922
soma enriched genes
n=5673 n=2553 n=13539 n=1305 n=19288 n=1492 n=9663 n=1909 n=7793 n=716
B
Differential expression between wild type
C. elegans (N2) and
a DCC mutant (e428), mRNA-seq
log2 fold change (e428/N2)
I II III IV V X
n=1980 n=1984 n=1943 n=1986 n=2421 n=1858
C
Distribution of |log2(hermaphrodite/male)|
mRNA-seq, O. tipulae
count
log2=3
germline enriched genes
chromosome
I
II
III
IV
V
X
|log2(hermaphrodite/male|

D
Wilcoxon rank sum test, one vs rest
-log10(p-value)
inf
B. malayi
P. pacificus
C. elegans (N2)
C. elegans (e428)
O. tipulae
H. contortus
chromosome
I
II
III
IV
V
X

FIGURE 5

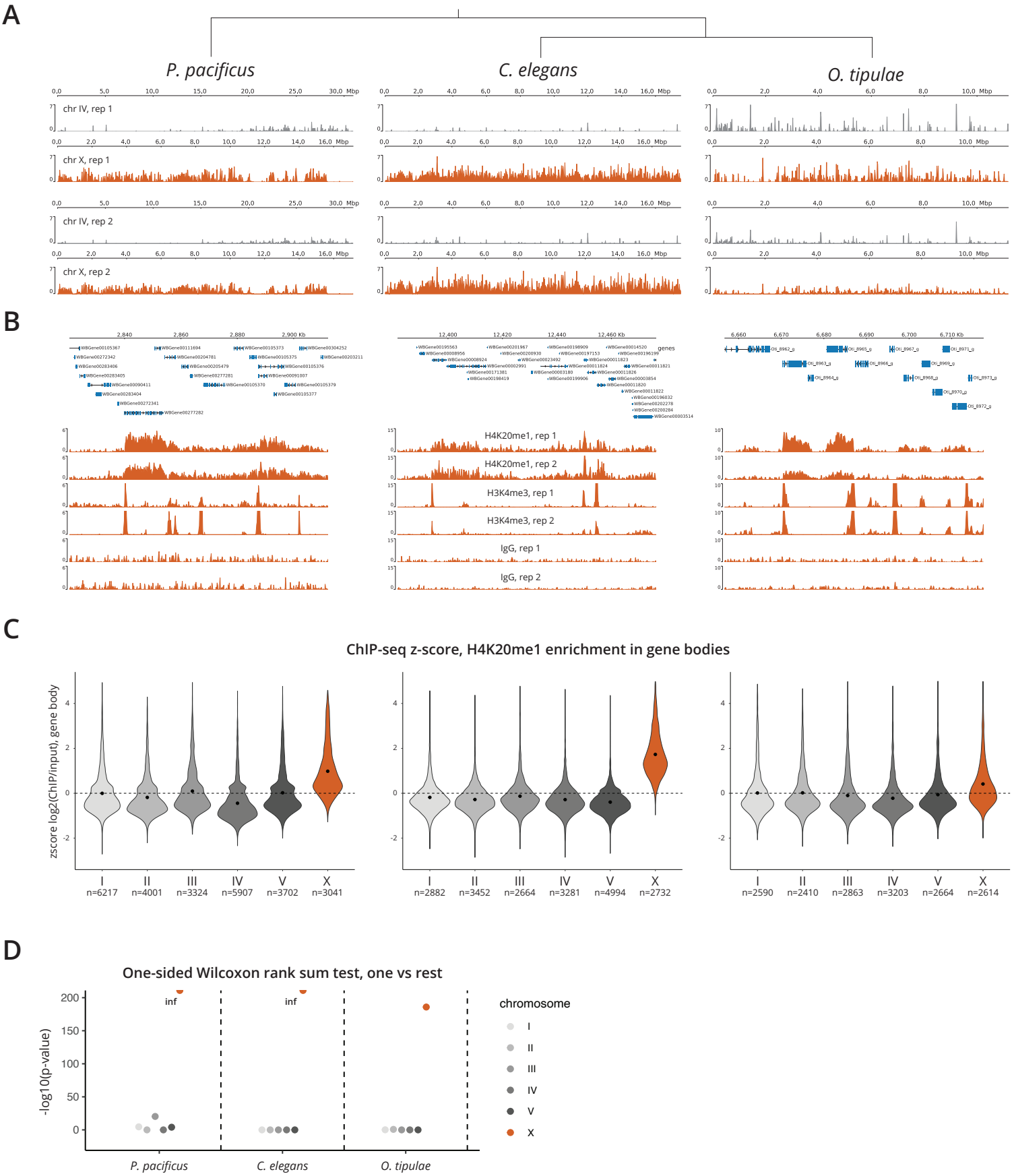

A
P. pacificus
C. elegans
O. tipulae
chr IV, rep 1
chr X, rep 1
chr IV, rep 2
chr X, rep 2
B
genes
H4K20me1, rep 1
H4K20me1, rep 2
H3K4me3, rep 1
H3K4me3, rep 2
IgG, rep 1
IgG, rep 2
C
ChIP-seq z-score, H4K20me1 enrichment in gene bodies
zscore log2(ChIP/input), gene body
I n=6217
II n=4001
III n=3324
IV n=5907
V n=3702
X n=3041
I n=2882
II n=3452
III n=2664
IV n=3281
V n=4994
X n=2732
I n=2590
II n=2410
III n=2863
IV n=3203
V n=2664
X n=2614
D
One-sided Wilcoxon rank sum test, one vs rest
-log10(p-value)
inf
inf
P. pacificus
C. elegans
O. tipulae
chromosome
I
II
III
IV
V
X

FIGURE 6

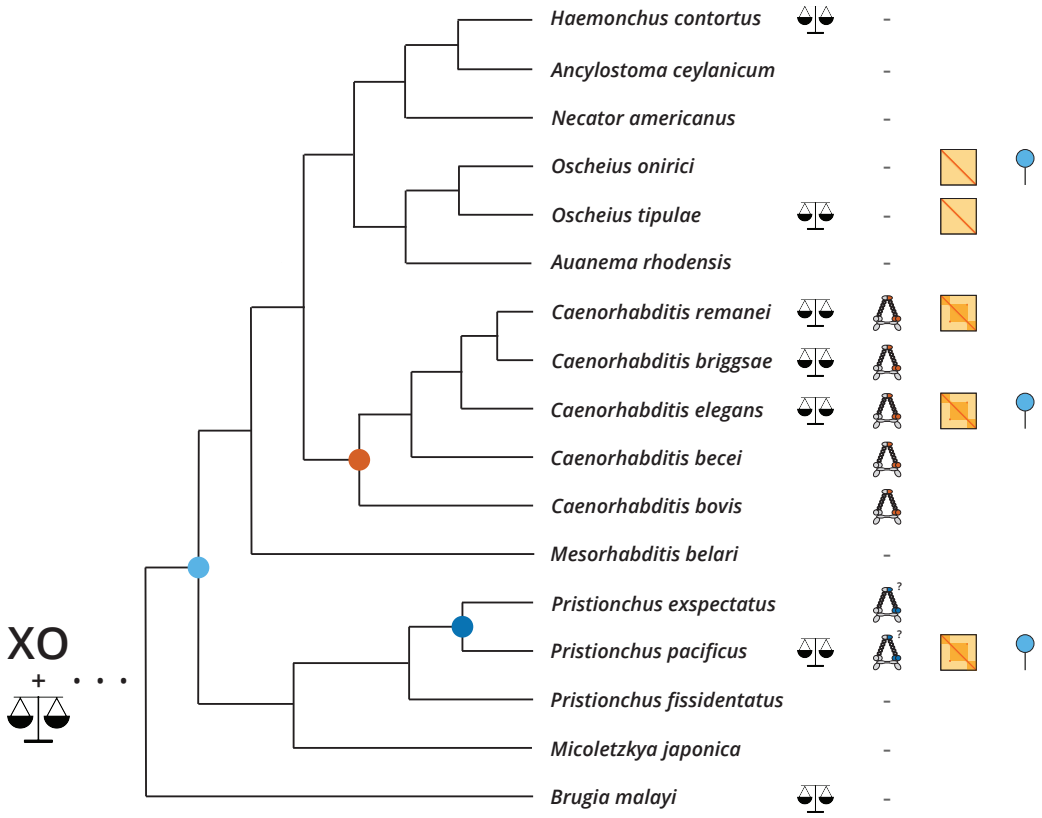

Haemonchus contortus
Ancylostoma ceylanicum
Necator americanus
Oscheius onirici
Oscheius tipulae
Auanema rhodensis
Caenorhabditis remanei
Caenorhabditis briggsae
Caenorhabditis elegans
Caenorhabditis becei
Caenorhabditis bovis
Mesorhabditis belari
Pristionchus exspectatus
Pristionchus pacificus
Pristionchus fissidentatus
Micoletzkya japonica
Brugia malayi
XO
+ . . .

FIGURE S1

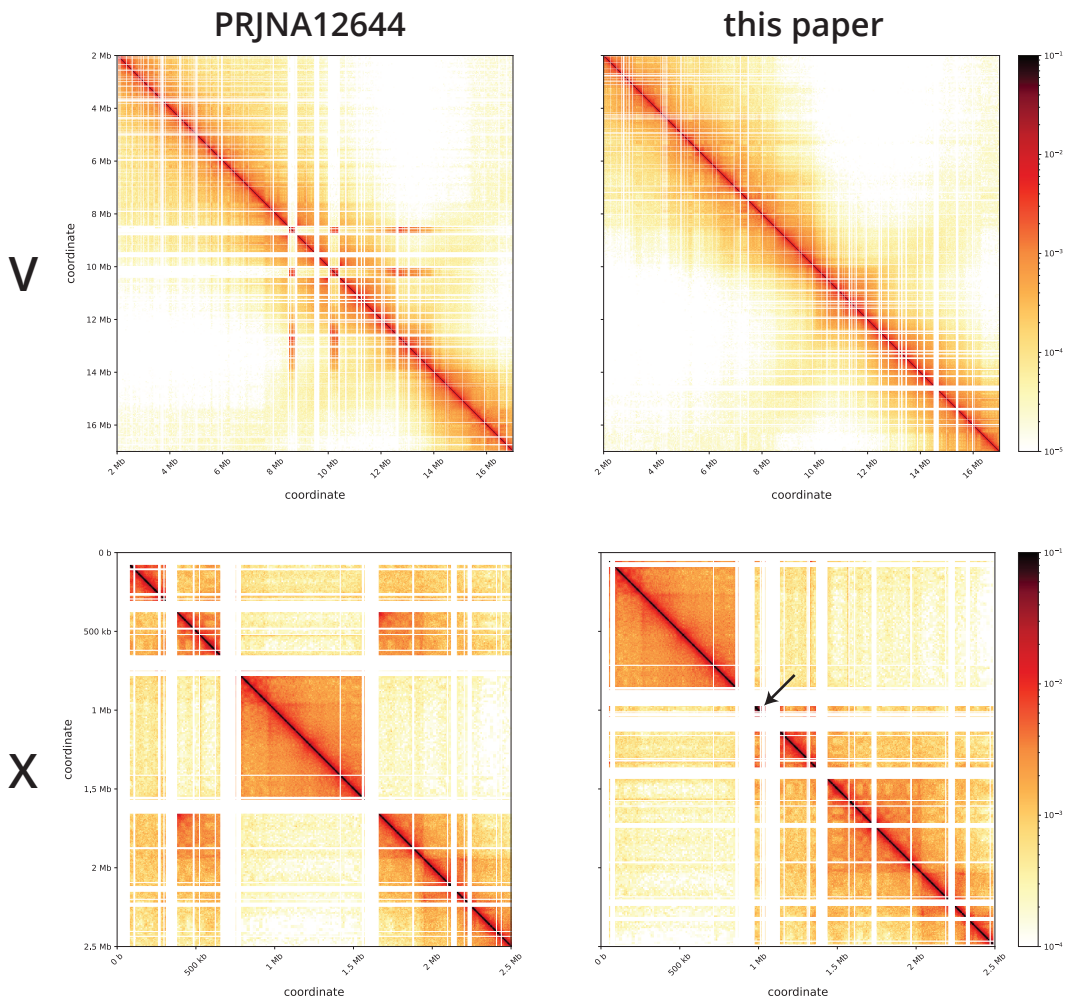

PRJNA12644
this paper
V
X
coordinate
coordinate
2 Mb
4 Mb
6 Mb
8 Mb
10 Mb
12 Mb
14 Mb
16 Mb
0 b
500 kb
1 Mb
1,5 Mb
2 Mb
2,5 Mb

FIGURE S2

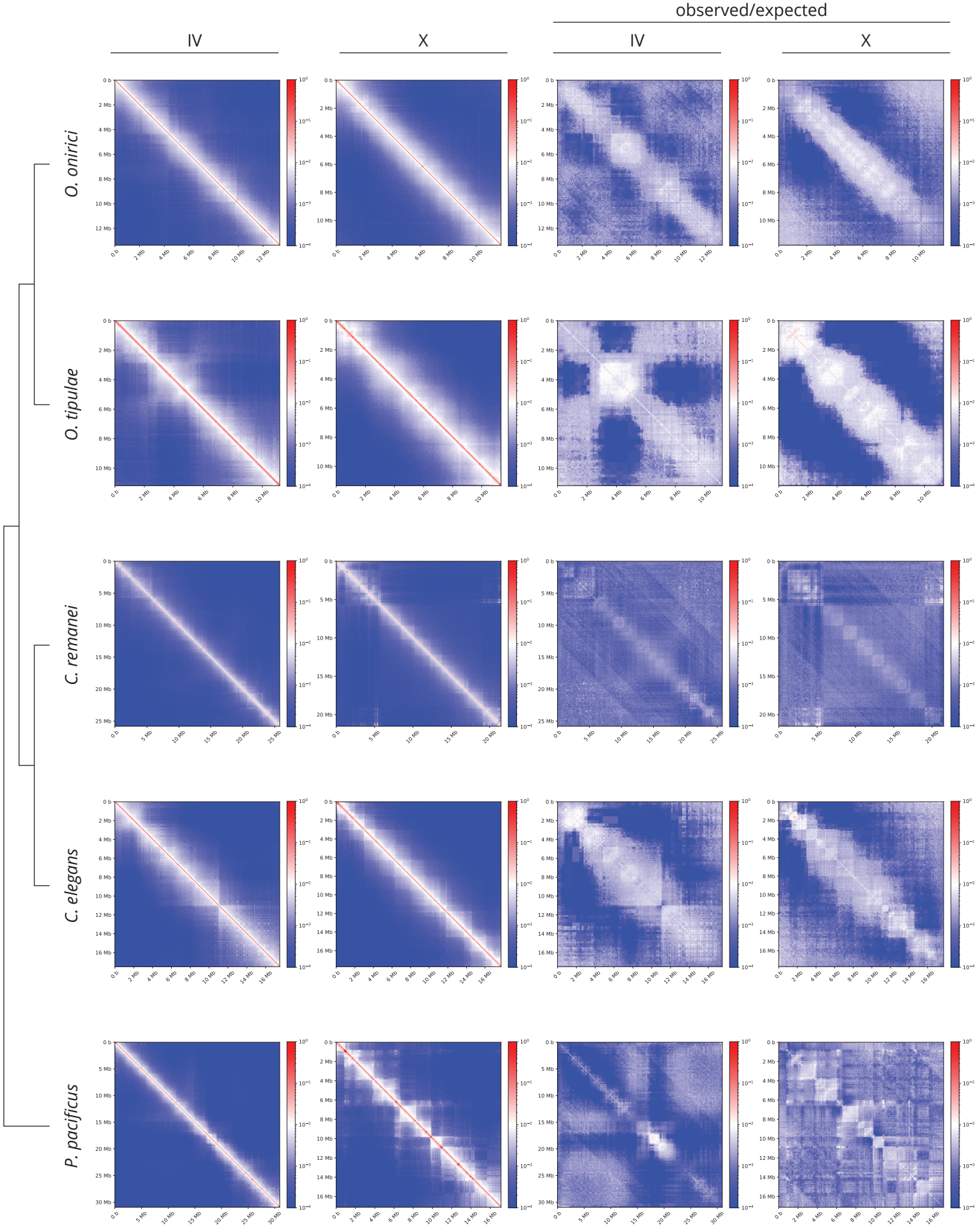

observed/expected
IV
X
IV
X
O. onirici
O. tipulae
C. remanei
C. elegans
P. pacificus

FIGURE S3

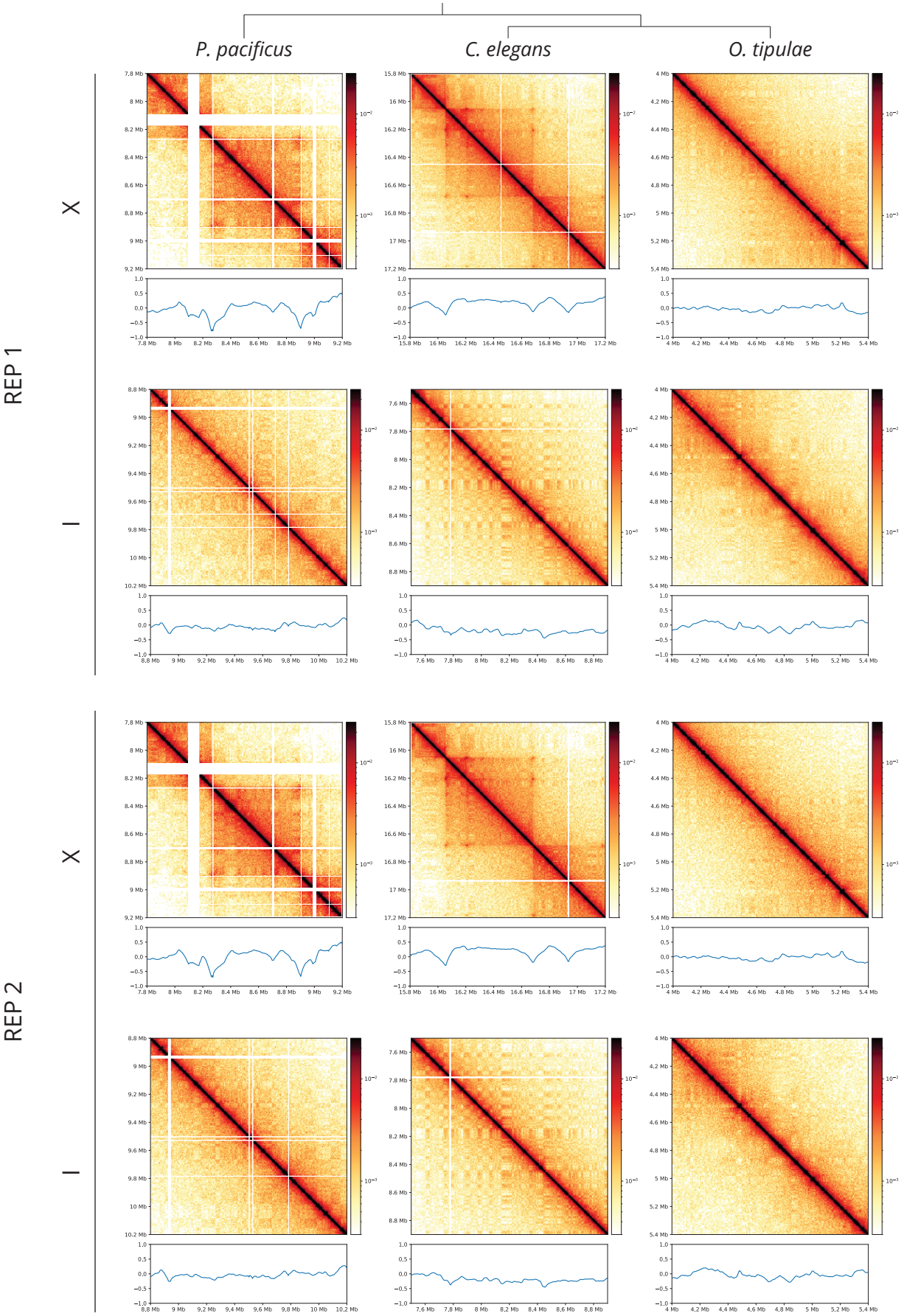


FIGURE S4

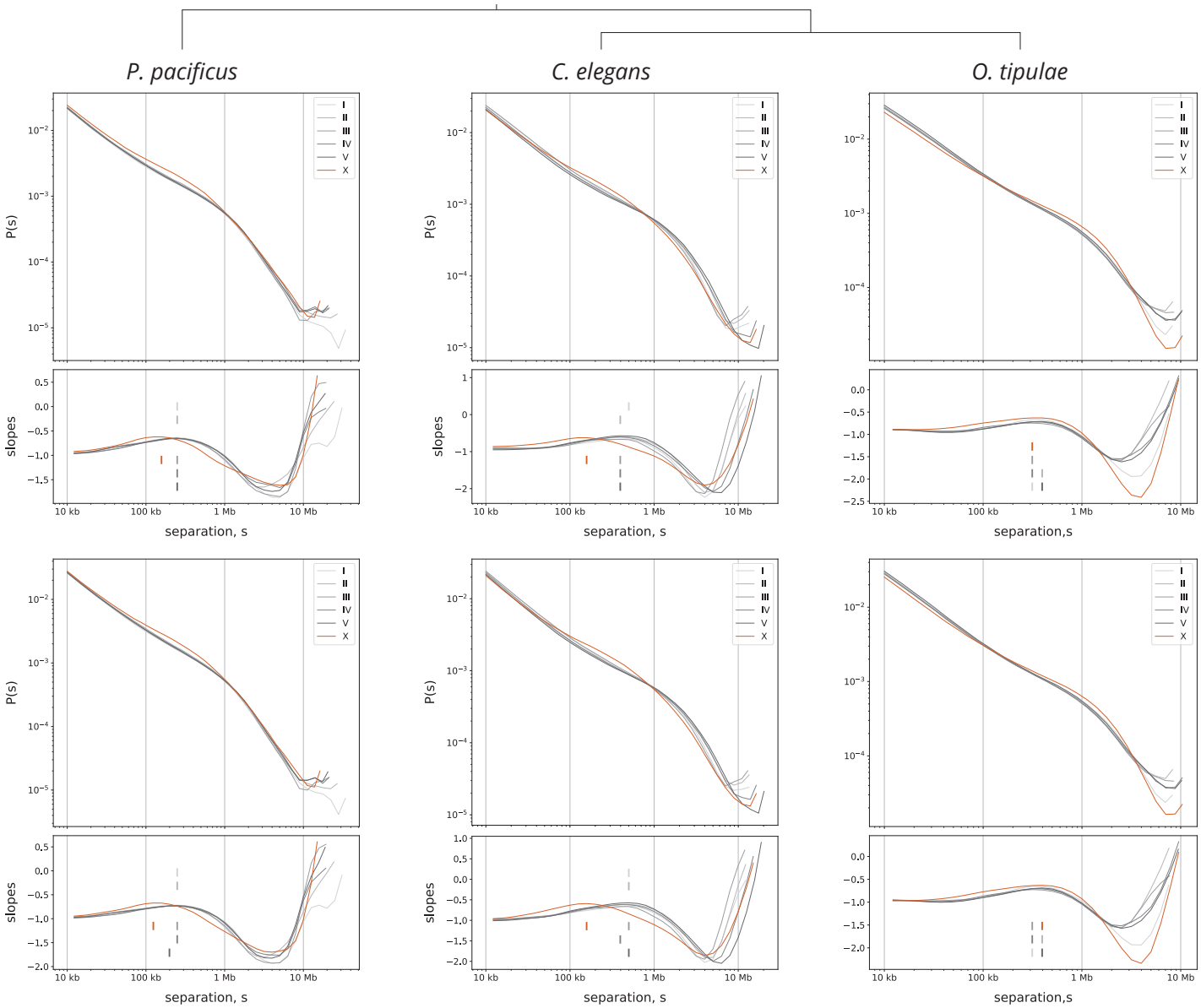

P. pacificus
C. elegans
O. tipulae
P(s)
slopes
separation, s
separation,s
10 kb
100 kb
1 Mb
10 Mb
I
II
III
IV
V
X

FIGURE S5

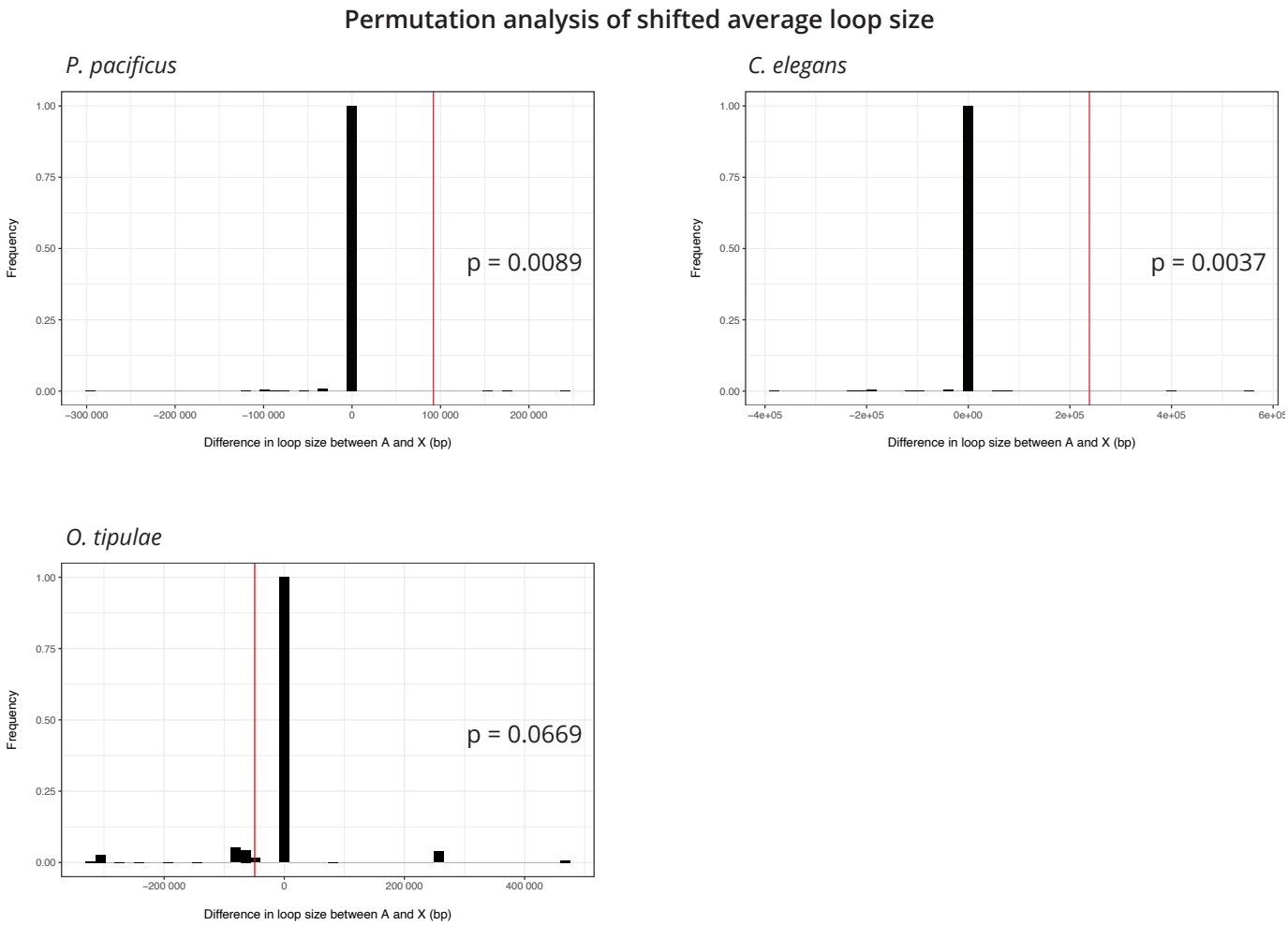
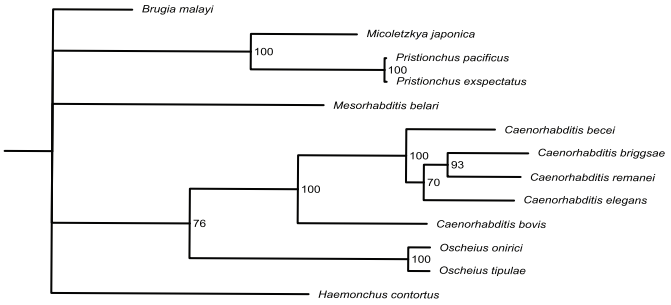
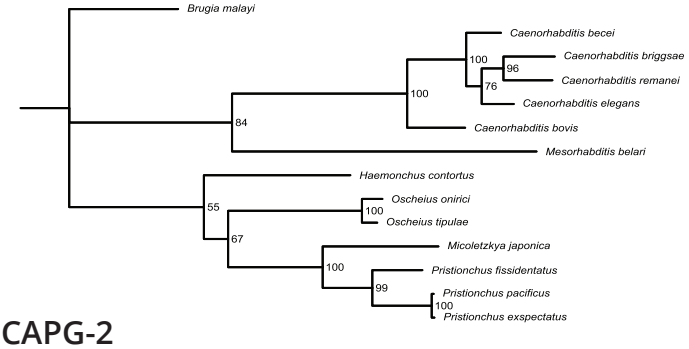

Permutation analysis of shifted average loop size
P. pacificus
p = 0.0089
Frequency
1.00
0.75
0.50
0.25
0.00
−300 000
−200 000
−100 000
0
100 000
200 000
Difference in loop size between A and X (bp)
C. elegans
p = 0.0037
Frequency
1.00
0.75
0.50
0.25
0.00
−4e+05
−2e+05
0e+00
2e+05
4e+05
6e+05
Difference in loop size between A and X (bp)
O. tipulae
p = 0.0669
Frequency
1.00
0.75
0.50
0.25
0.00
−200 000
0
200 000
400 000
Difference in loop size between A and X (bp)

FIGURE S6

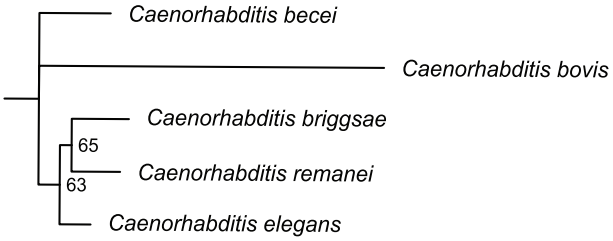
MIX-I



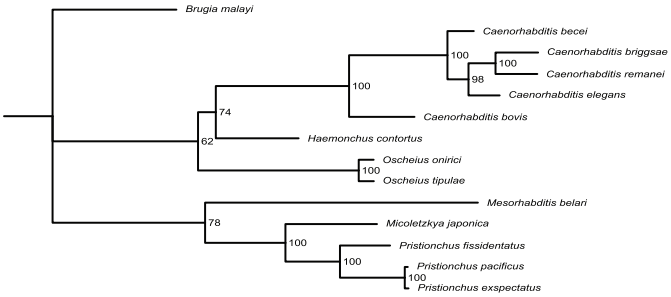
KLE-2



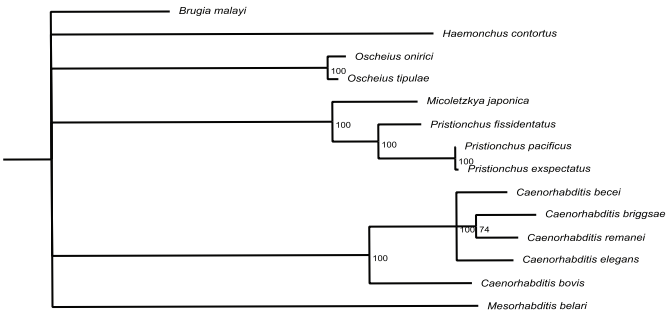
DPY-26



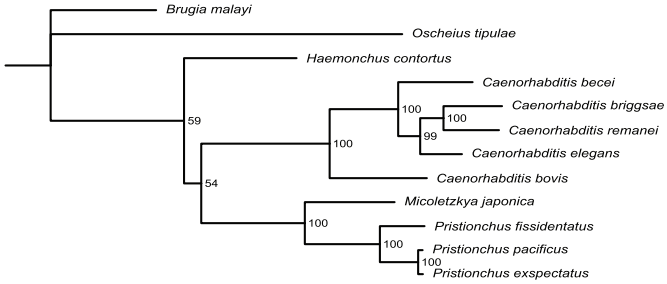
CAPG-2



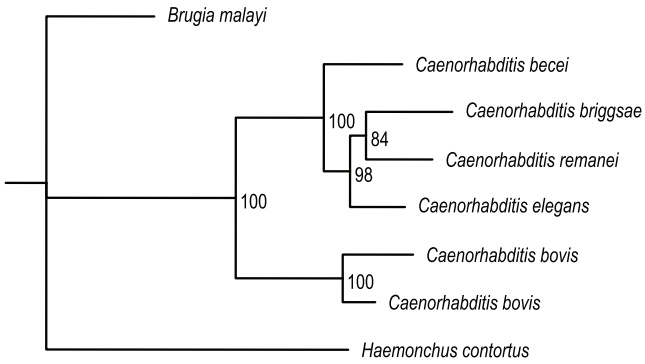
DPY-28



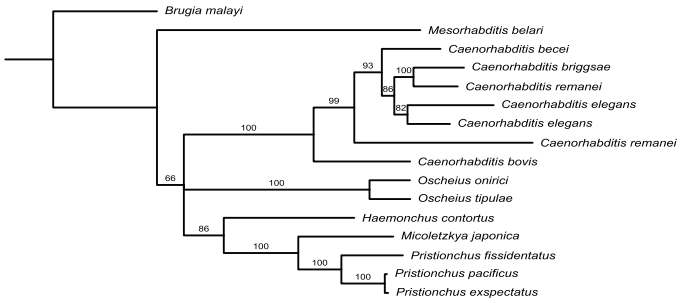
SMC-5



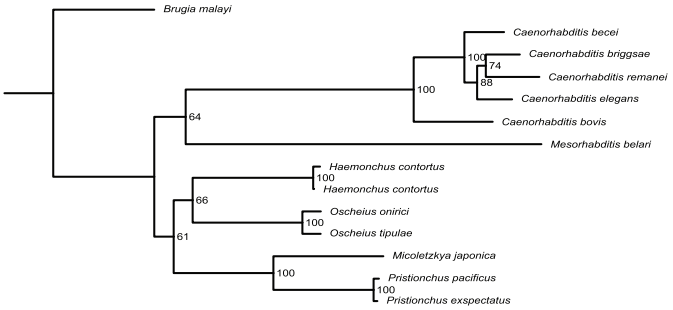
CAPG-1



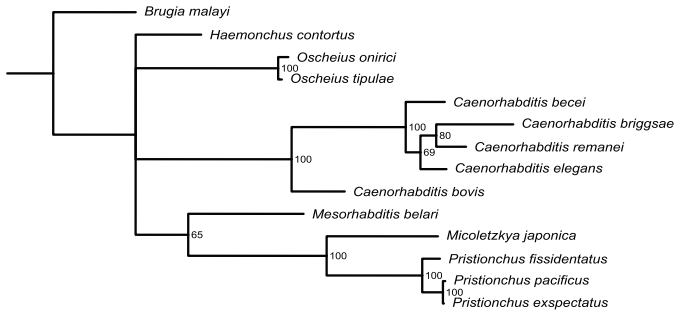
SMC-6



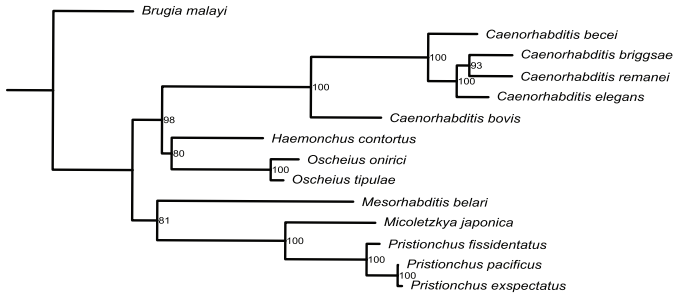
HCP-6



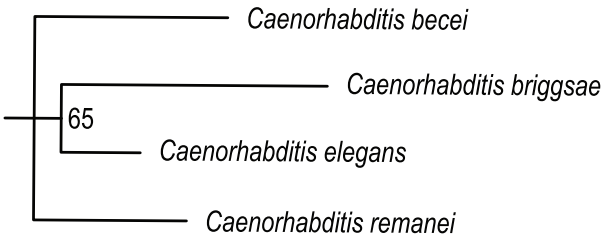
SMC-1



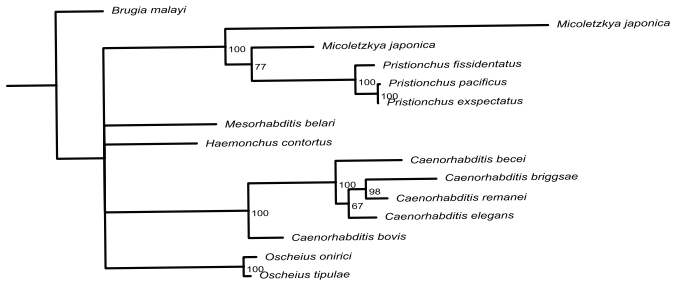
SCC-3



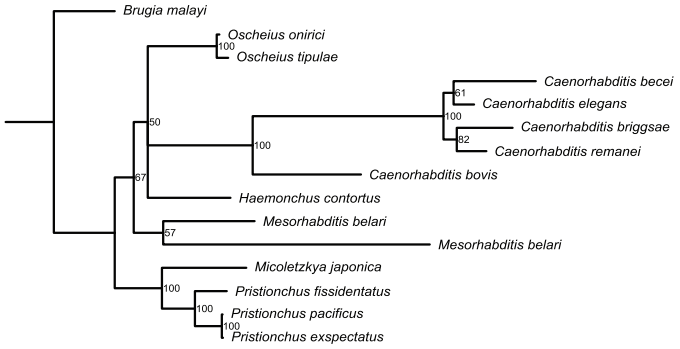
REC-8



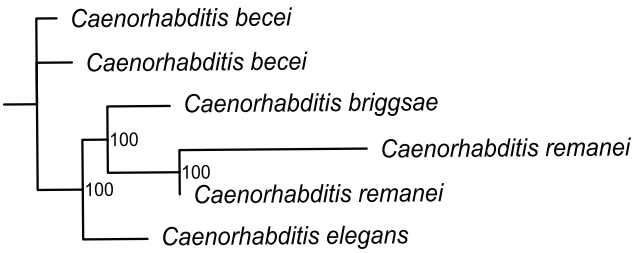
SMC-3



COH-1



COH-2



COH-3/4

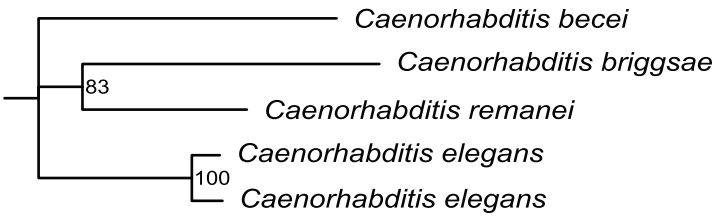


FIGURE S7

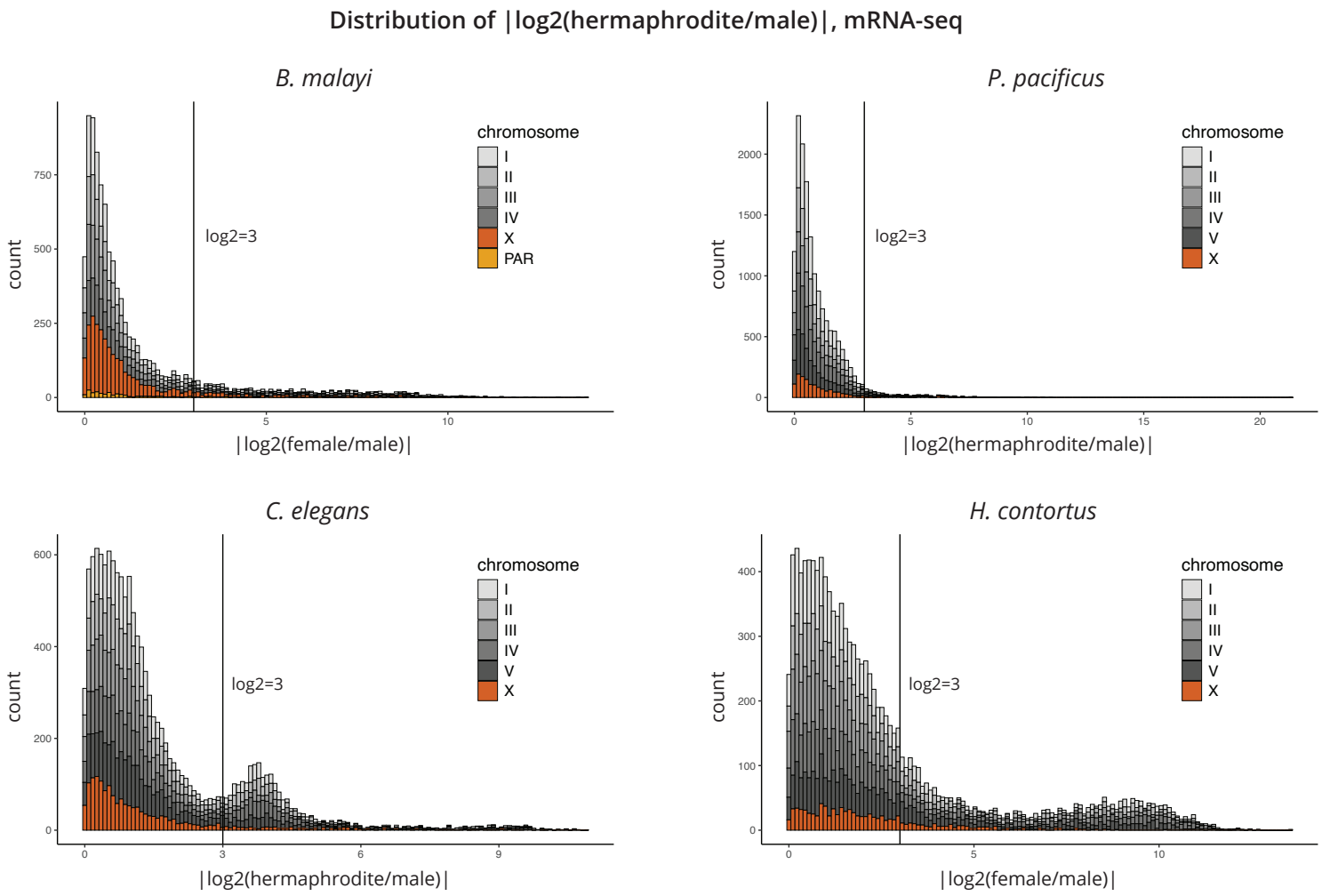

Distribution of |log2(hermaphrodite/male)|, mRNA-seq
B. malayi
P. pacificus
C. elegans
H. contortus
chromosome
I
II
III
IV
V
X
PAR
log2=3
count
|log2(female/male)|
|log2(hermaphrodite/male)|

FIGURE S8

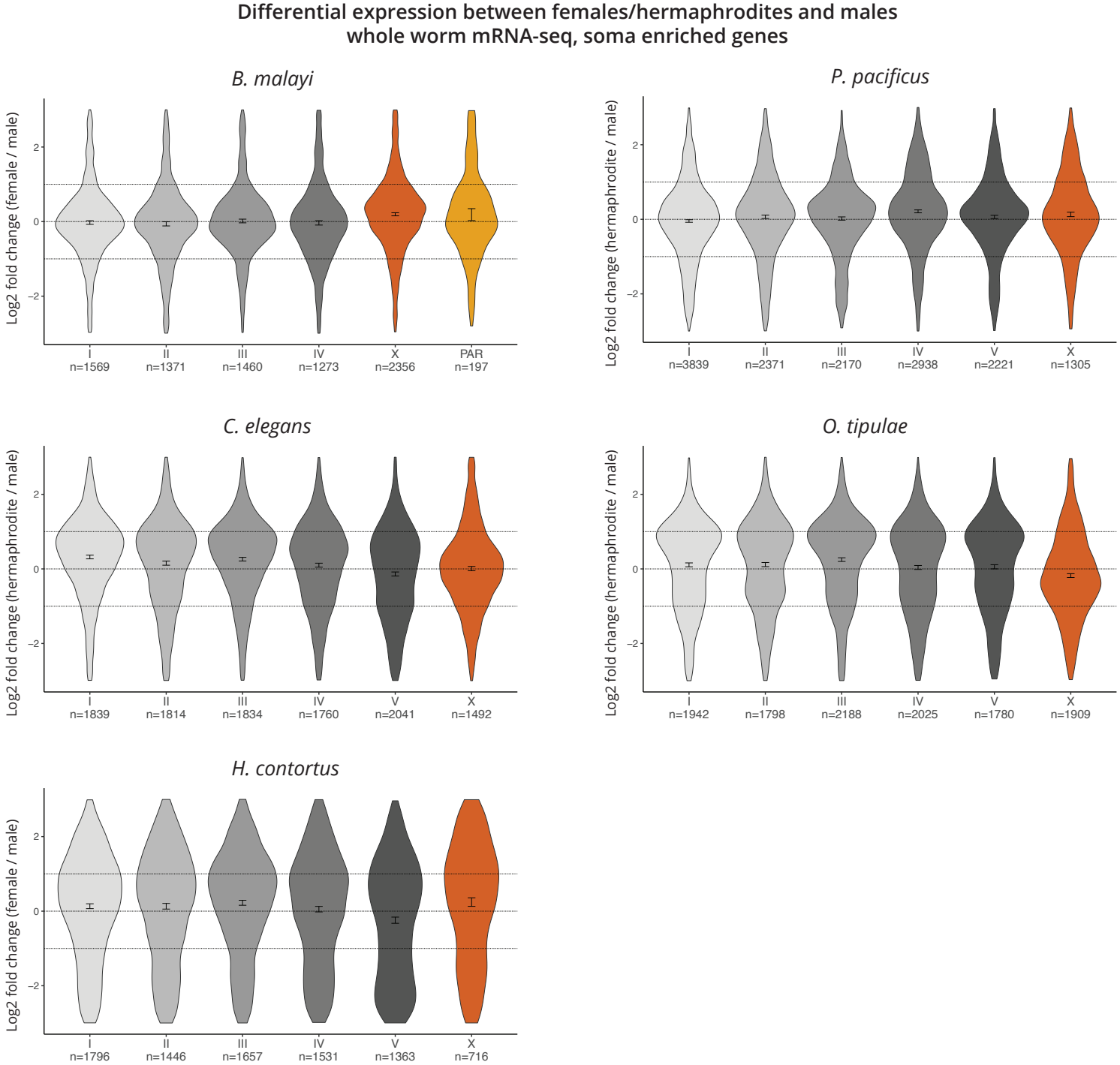
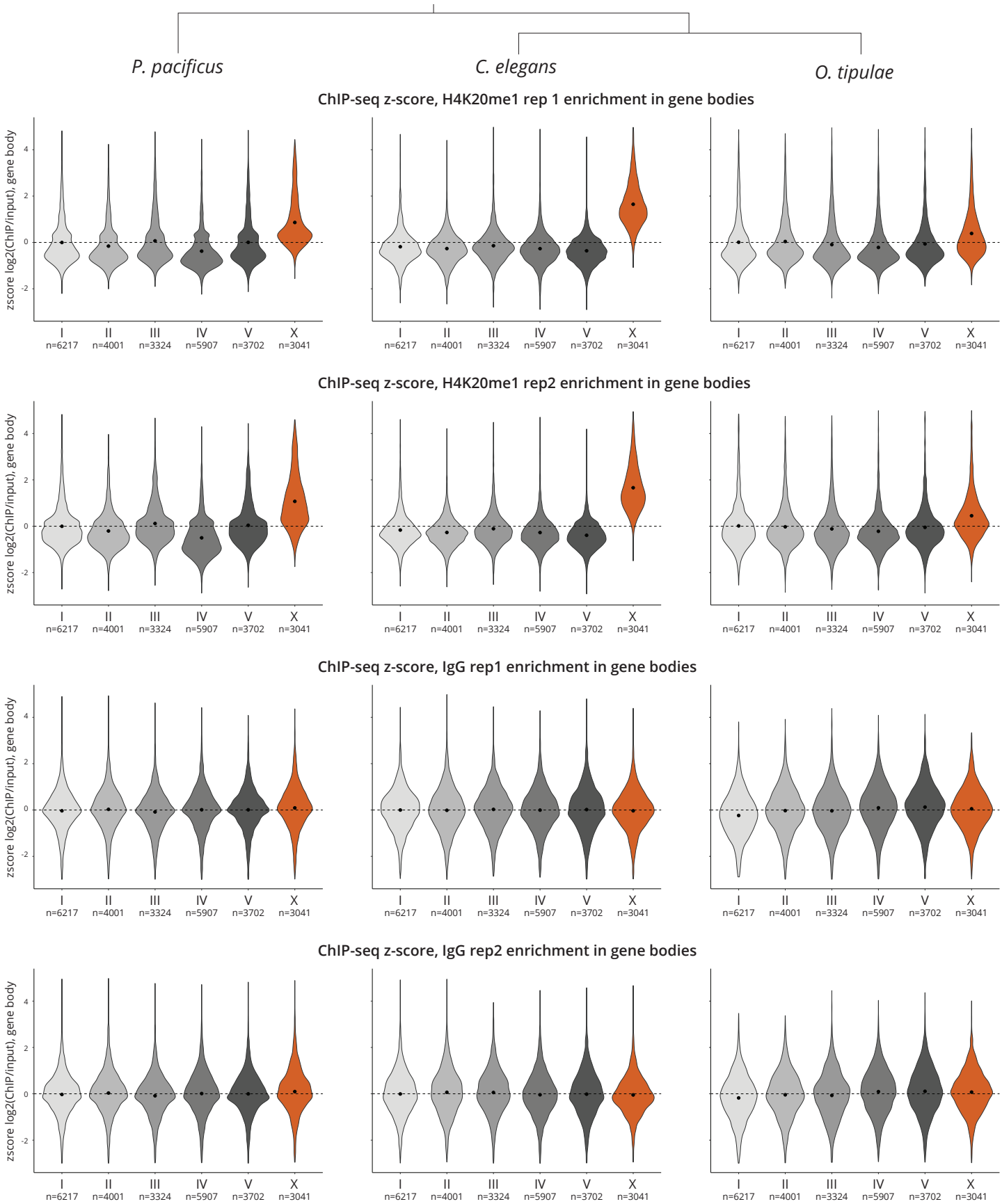

Differential expression between females/hermaphrodites and males
whole worm mRNA-seq, soma enriched genes
B. malayi
Log2 fold change (female / male)
2
0
−2
I n=1569
II n=1371
III n=1460
IV n=1273
X n=2356
PAR n=197
P. pacificus
Log2 fold change (hermaphrodite / male)
I n=3839
II n=2371
III n=2170
IV n=2938
V n=2221
X n=1305
C. elegans
Log2 fold change (hermaphrodite / male)
I n=1839
II n=1814
III n=1834
IV n=1760
V n=2041
X n=1492
O. tipulae
Log2 fold change (hermaphrodite / male)
I n=1942
II n=1798
III n=2188
IV n=2025
V n=1780
X n=1909
H. contortus
Log2 fold change (female / male)
I n=1796
II n=1446
III n=1657
IV n=1531
V n=1363
X n=716

FIGURE S9


P. pacificus
C. elegans
O. tipulae
ChIP-seq z-score, H4K20me1 rep 1 enrichment in gene bodies
ChIP-seq z-score, H4K20me1 rep2 enrichment in gene bodies
ChIP-seq z-score, IgG rep1 enrichment in gene bodies
ChIP-seq z-score, IgG rep2 enrichment in gene bodies
zscore log2(ChIP/input), gene body
4
2
0
-2
I n=6217
II n=4001
III n=3324
IV n=5907
V n=3702
X n=3041