**SUPPLEMENTARY INFORMATION**

**Single worm long-read sequencing reveals genome diversity in free-living nematodes**

Yi-Chien Lee1,2,3, Huei-Mien Ke4, Yu-Ching Liu1, Hsin-Han Lee1, Min-Chen Wang5, Yung-Che Tseng5, Taisei Kikuchi6 and Isheng Jason Tsai1,2,\*

1Biodiversity Research Center, Academia Sinica, Taipei 115, Taiwan

2Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National Taiwan Normal University, Taipei, Taiwan

3Department of Life Science, National Taiwan Normal University, 116 Wenshan, Taipei, Taiwan

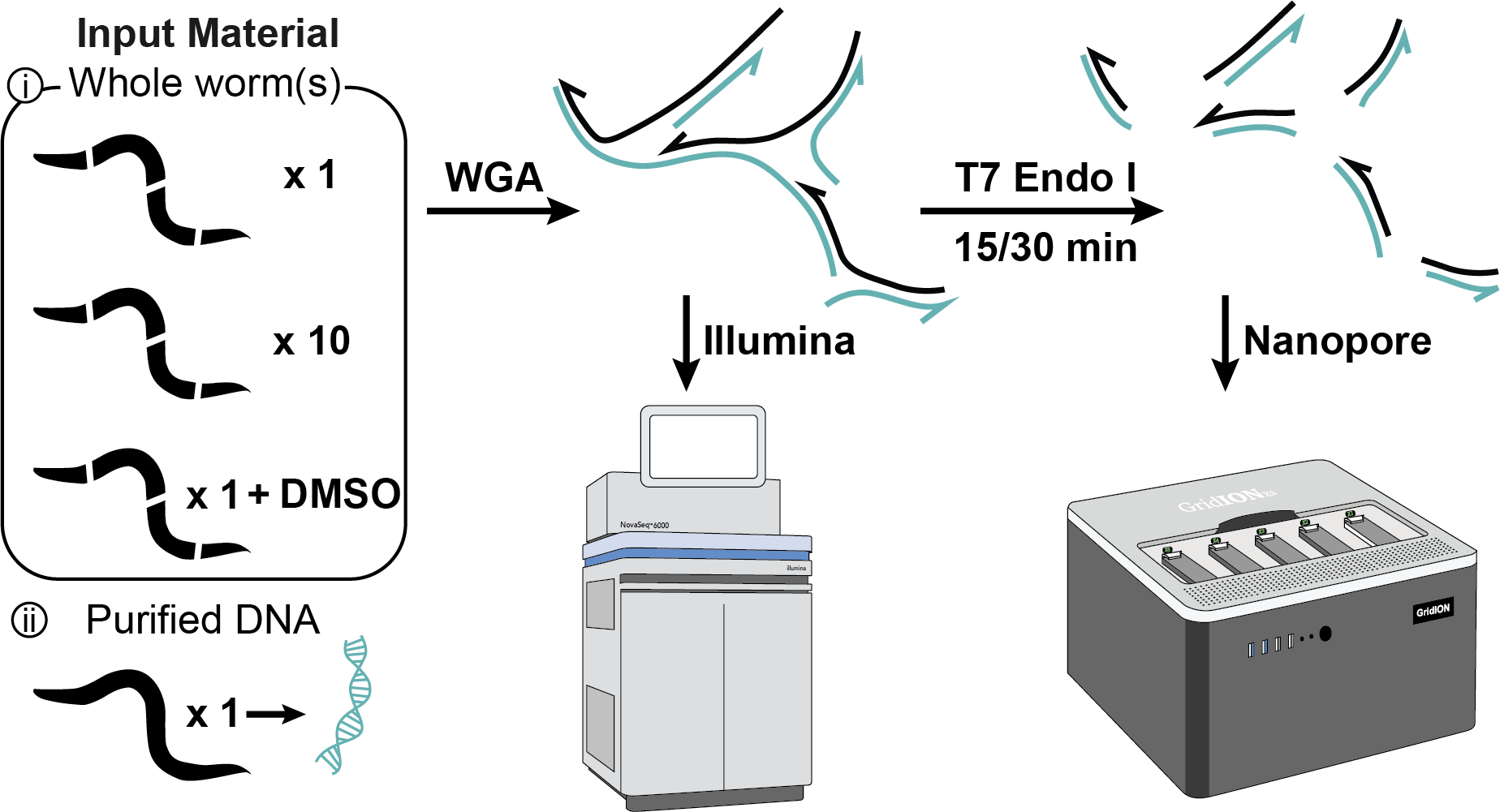
4Department of Microbiology, Soochow University, Taipei, Taiwan

5Marine Research Station (MRS), Institute of Cellular and Organismic Biology, Academia Sinica, I-Lan County, Taiwan

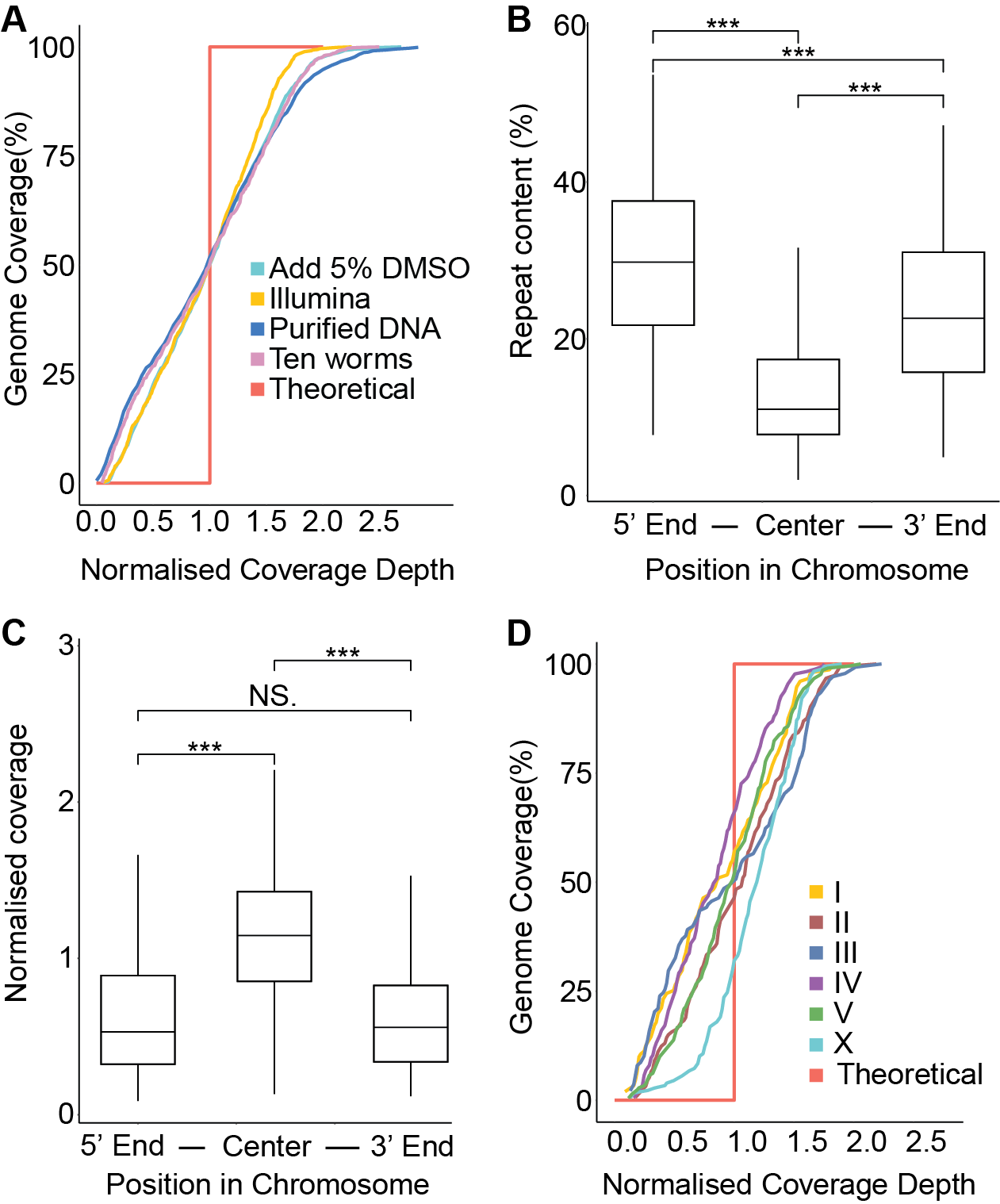
6Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, 277-8562, Japan

\* Correspondence to: Isheng Jason Tsai, Email: [ijtsai@gate.sinica.edu.tw](mailto:ijtsai@gate.sinica.edu.tw)

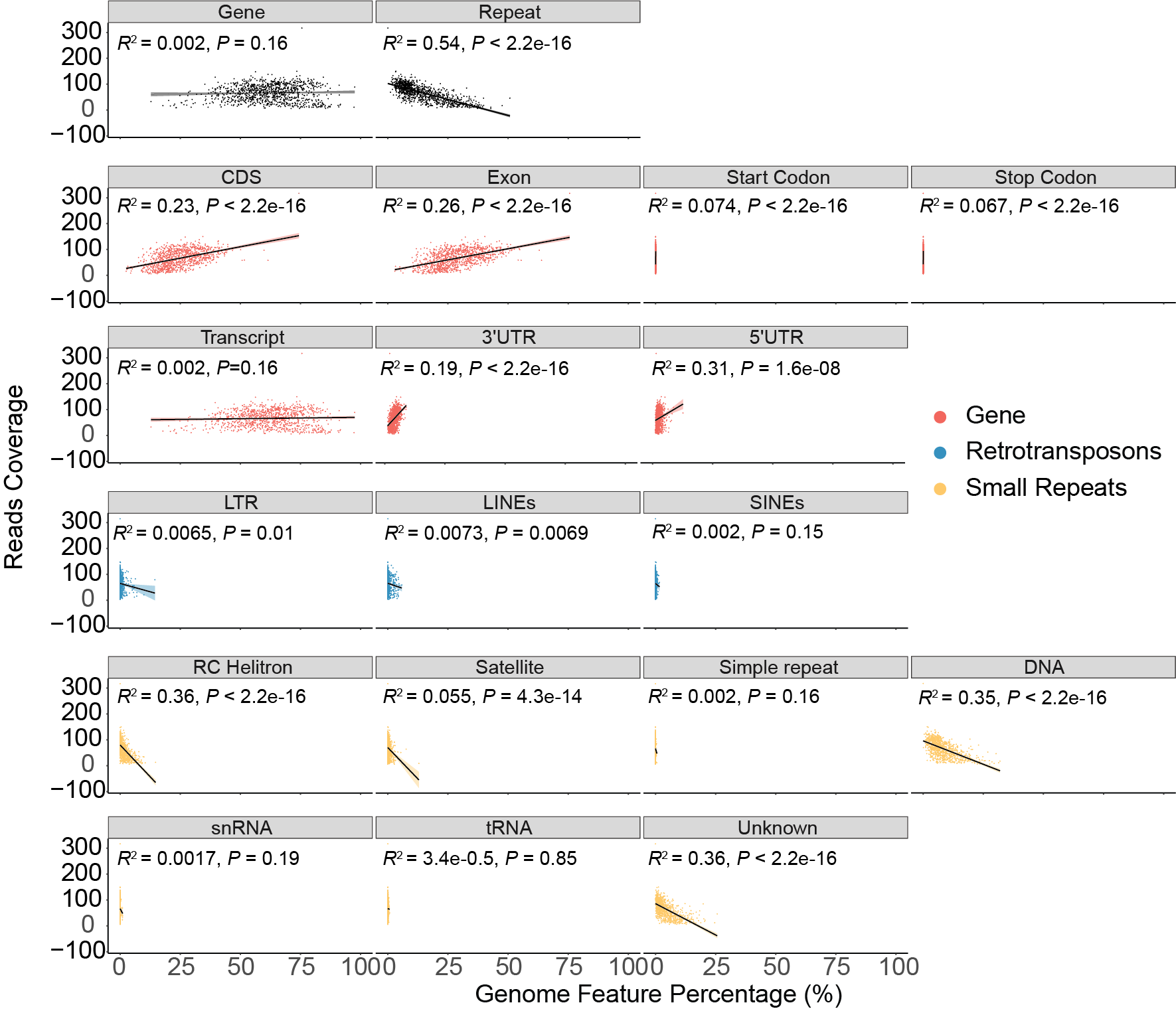
**Figure S1. Schematic diagram showing the strategies we used to optimise the whole genome amplification (WGA) step.** Four types of input material were used: single worm, ten worms pooled, single worm with 5% DMSO, and purified gDNA from single worm. After WGA, templates were sequenced with an Illumina sequencer. Templates digested with different T7 end I treatment times were sequenced with an Oxford Nanopore platform.



**Figure S2. Repeat content and sequencing coverage in *C. elegans*   
A.** Cumulative Illumina read genome coverage versus genome wide median in *C. elegans*. **B.** Proportion of repeat content in 5’end, center and 3’end region of *C. elegans* six chromosomes. **C.** Normalised read coverage in 5’end, center and 3’end of six chromosomes. \*: *P* 0.05 \*\*\*: *P* 0.001, NS.:*P* > 0.05. **D.** Cumulative Nanopore read genome coverage versus genome wide median in six chromosomes of *C. elegans*.

****

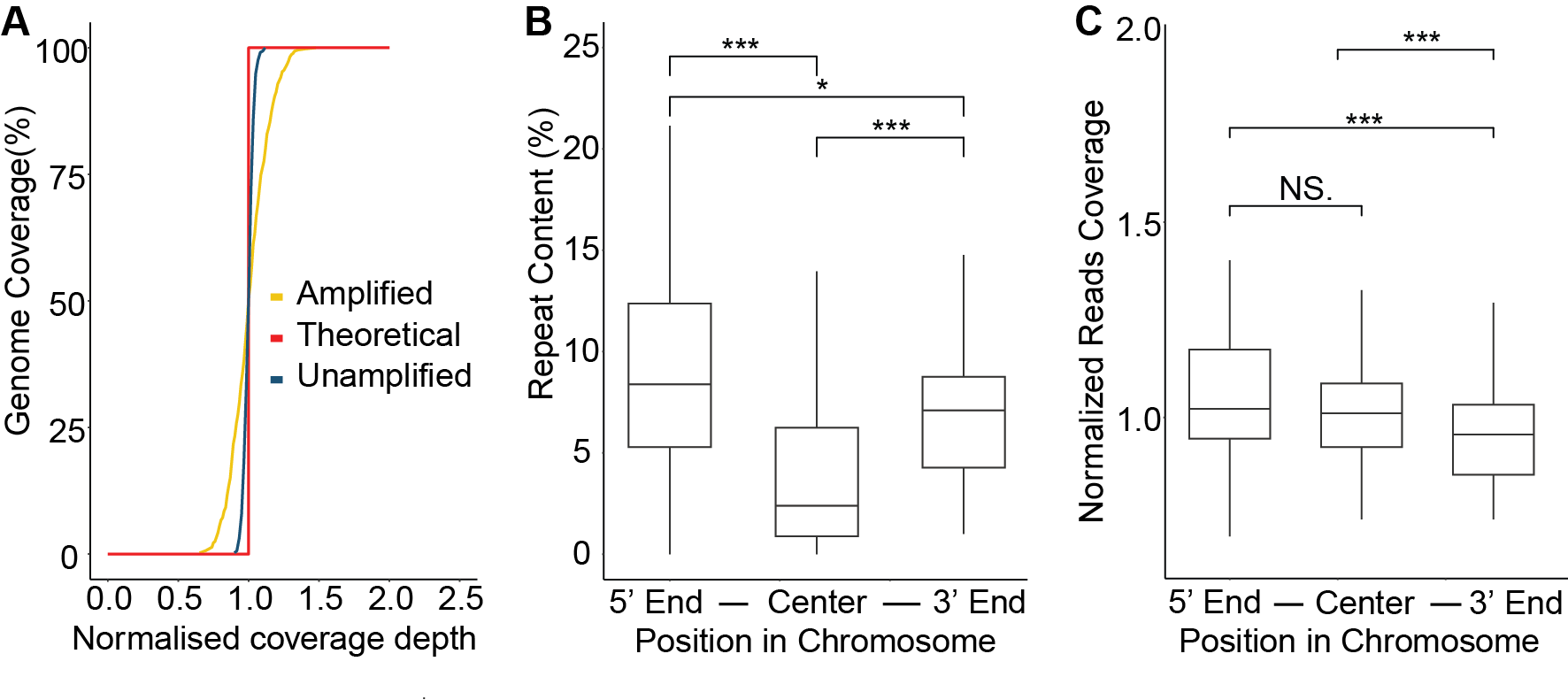
**Figure S3. Correlation between repeat content and sequence coverage in non-overlapping 100-kb windows.** Scatter plots showing read coverage of different genome features, including repeats and genes in amplified samples. Genome coverage and repetitive content (RC Helitron, and DNA transposons, and unknown repeats) were siginificantly negatively correlated.

****

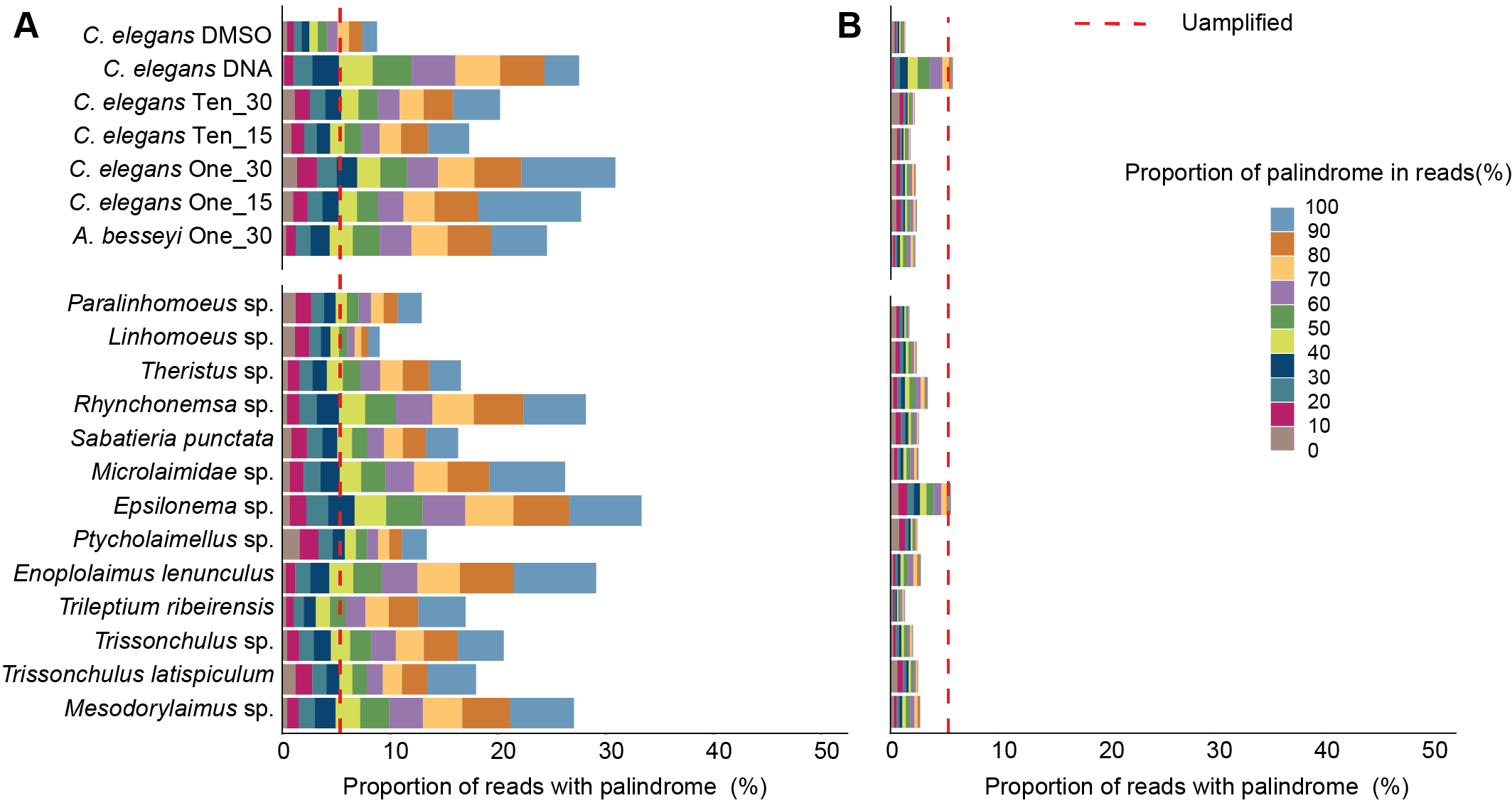
**Figure S4. Reads coverage across genome features.** The box plot shows the normalised reads coverage on each genome feature. RC Helitron had the lowest coverage. The numbers box represent the median coverage of each feature.

****

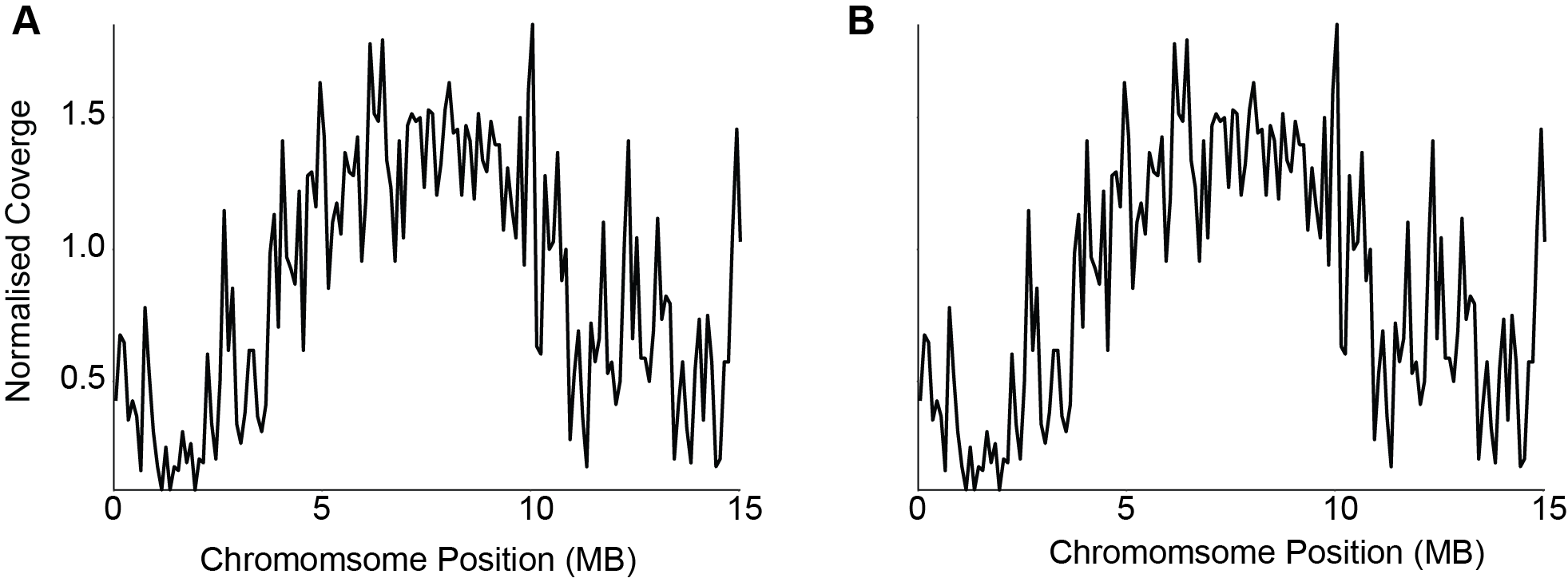
**Figure S5. Coverage uneveness of WGA on the *A. bessyi* genome. A.** Cumulative Nanopore reads genome coverage versus genomewide median in three chromosomes of *A. besseyi* genome. **B.** The repeat content in *A. besseyi* chromosomal 5’end, center and 3’end. **C**. Sequencing reads coverage in *A. besseyi* chromosomal 5’end, center and 3’end. \*: *P*  0.05 \*\*\*: *P* 0.001, NS.:*P* > 0.05.

****

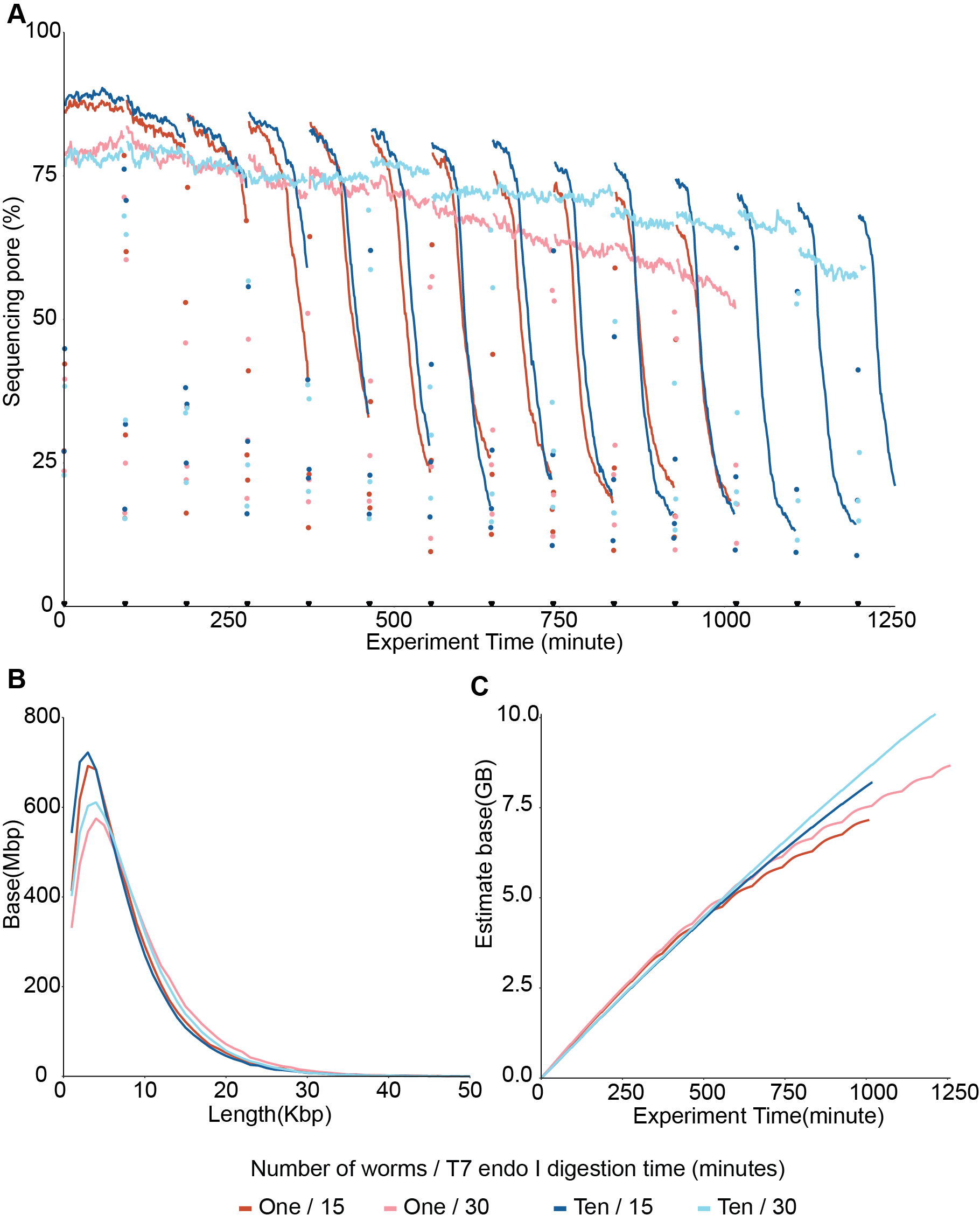
**Figure S6. Propotion of palindrome in reads. A.** Proportion of palindrome in the dataset prior of removing palindromes. **B.** Proportion of palindromes in dataset after correction. The dashed line represent the propotion of palindrome in the unamplfied *C. elegans* sample.

****

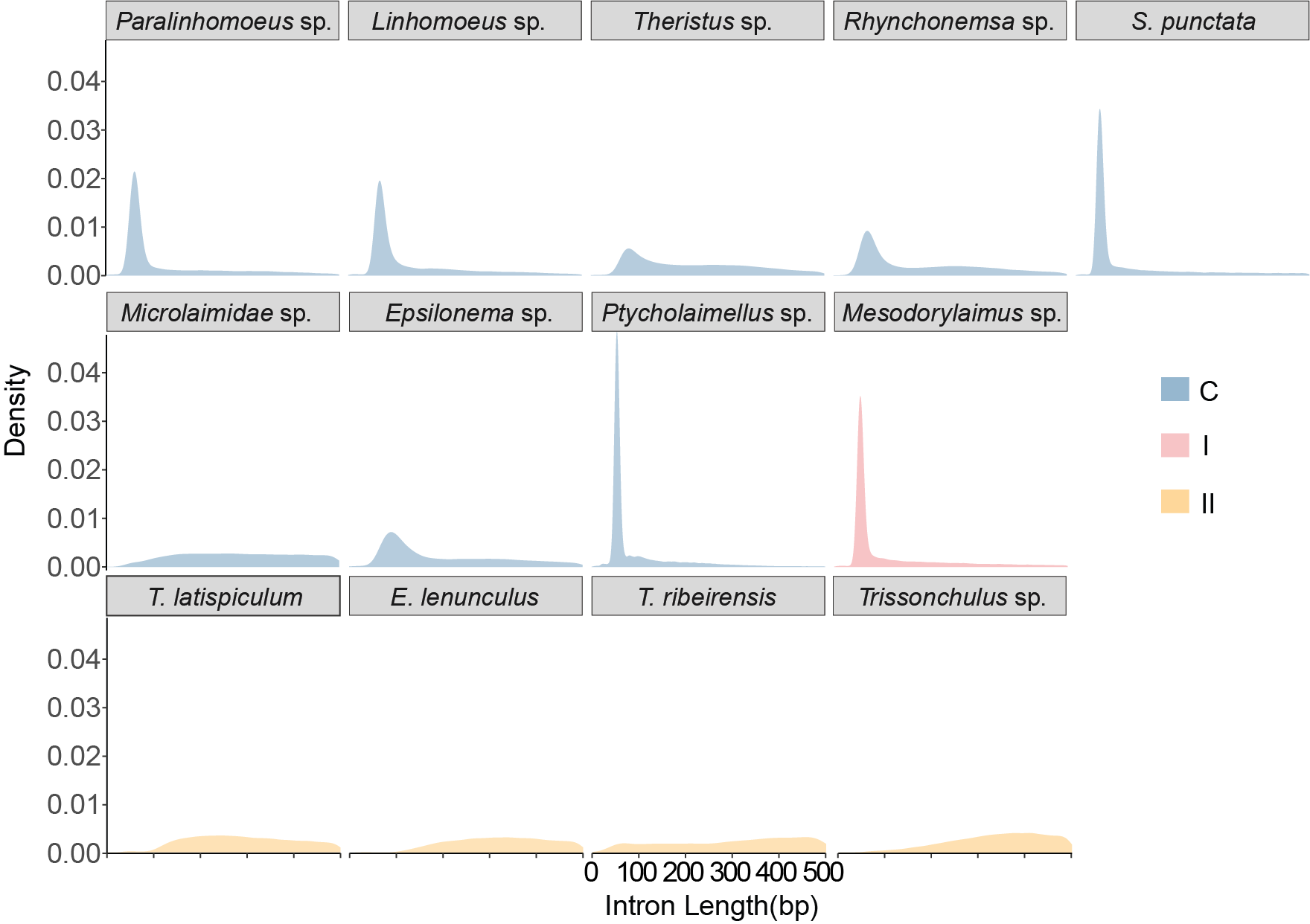
**Figure S7. The genome coverage of the palindrome reads paralleled that of non-palindrome reads.** **A.** Palindrome and **B.** non-palindrome reads coverage across the chromosome I of *C. elegans.*  (Pearson’s R2=1.0, *P* < 22e-16)

****

**Figure S8. Effect of T7 digest times on the Oxford Nanopore sequencing platform. A.** Number of sequencing pores over sequencing time. The triangles represent the time of mux scan. **B.** Reads distribution when subset 6.6 Gb data from each sample. **C.** Cumulative sequencing output (Gb) over time. Reads length distribution showed that with 15 minutes digestion, the reads length is generally shorter. This shows that longer digestion times indeed generate more linear templates and have fewer defects on sequencing.



**Figure S9. Distribution of intron length in nematodes.** Different colours denote different Nematoda clades.

****

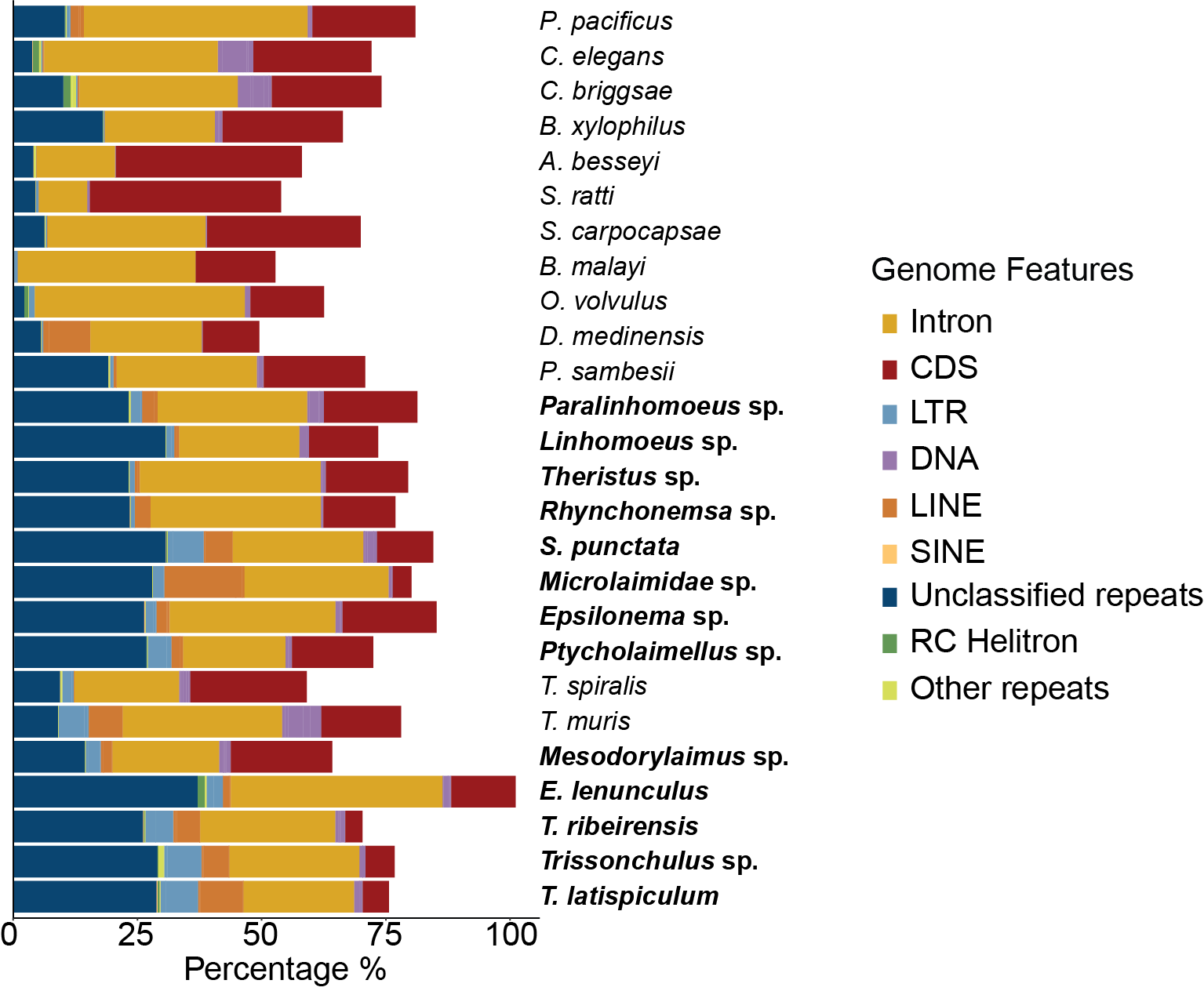
**Figure S10. Distribution of intron length (A) and numbers (B) in nematode proteomes.** Different colours denote different Nematoda clades.



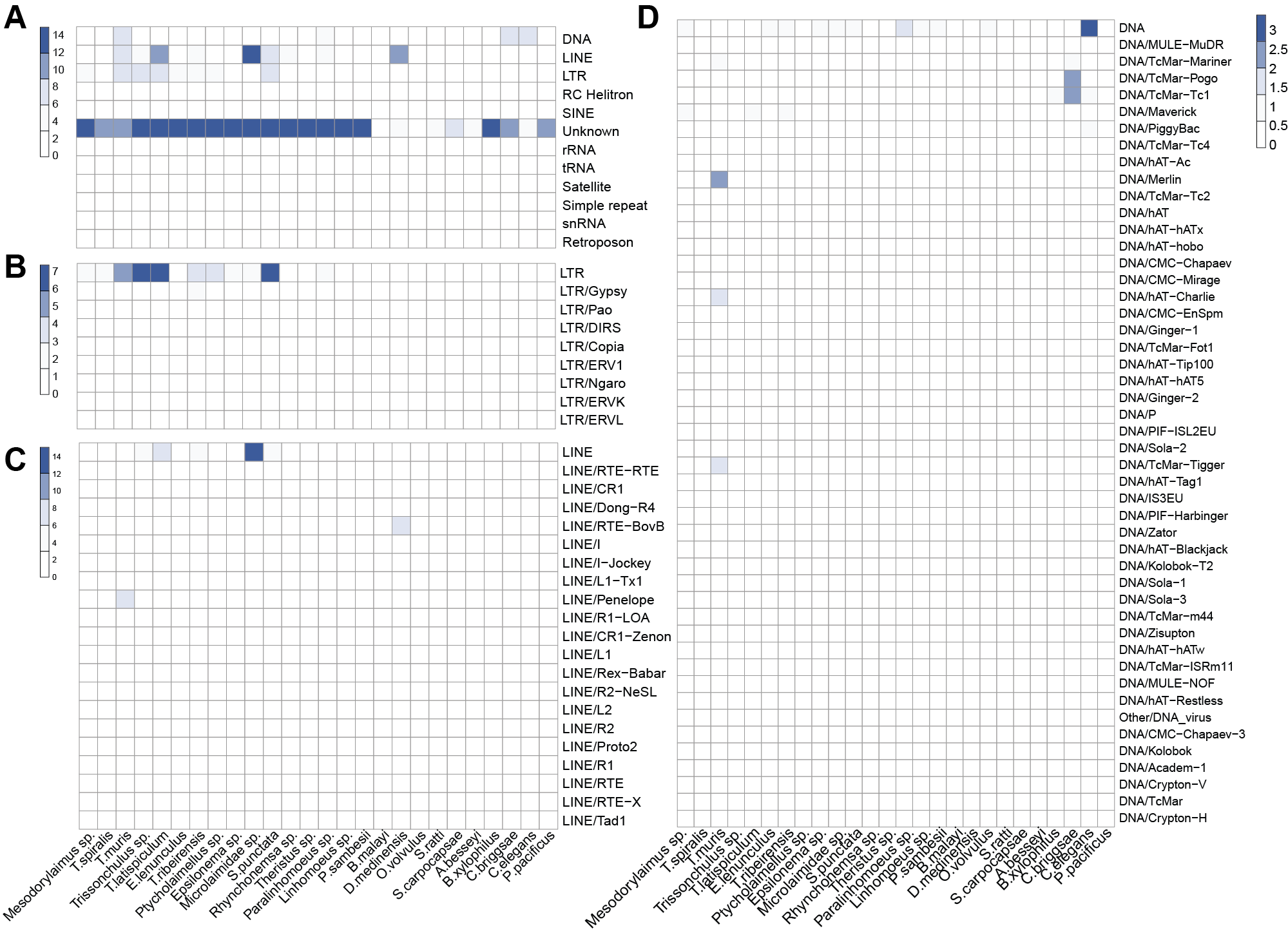
**Figure S11. Mitochondrial gene orders in nematodes.** The mitochondral gene order from representative published nematode mitogenomes in clades C and I, and nine species with near complete assembled mitogenomes in this study, are denoted by bold letters.



**Figure S12.** **Percentage of genome features in the nematode genomes**.   
Bold letters represent the nematode genome sequenced in this study.

****

**Figure S13. Repeat contents in nematode genomes. A.** The proportions of different repeat feature, coloured by percentages. **B-D.** Proportions of LTR, DNA, LINEs. The pattern of enrichment of DNA transposon, LTR, and LINE families appeared to be specific to each nematode.



**Figure S14. The Nematoda species tree.** The phylogeny tree was constructed with the proteomes of 26 nematode genomes, which included 13 published repersentive species and 13 assemblies (bold font) in this study. *Priapulus caudatus* is placed as the outgroup. The bootstrap support value of each node is 100.

