

Chromosome-Scale Assembly of Novel *Caenorhabditis* sp. 65 (strain JU4118)

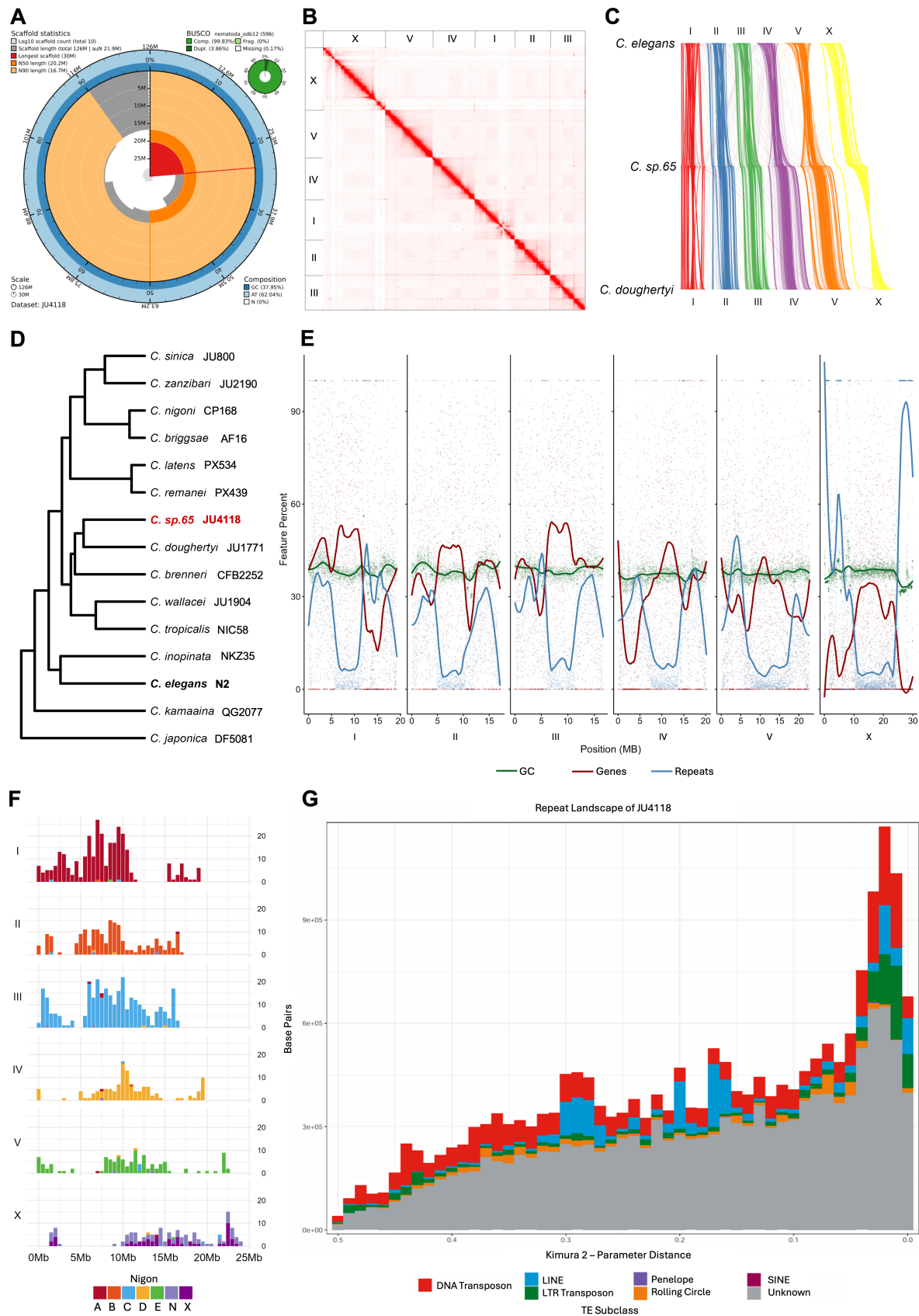
Jasbelle Sosa¹, Michelle A. McCauley¹, Victoria K. Eggers¹, Janna L. Fierst¹, Karolina Willicott^{1§}

¹Biological Sciences, Florida International University, Miami, FL, United States

[§]To whom correspondence should be addressed: kwillico@fiu.edu

Abstract

We present a chromosome-scale assembly of strain JU4118, an inbred line of *Caenorhabditis* sp. 65. This species is a wild isolate that was sampled in Da Lat, Vietnam by the Marie-Anne Félix lab in December 2019. This strain has a genome size of 126 Mb, with a GC content of 37.95 %. Comparisons of *C. sp. 65* with *C. elegans* show conserved broad-scale chromosomal organization, but finer-scale variation which may be clade-specific, such as expansion of the X chromosome and increased repeat content. By assembling and annotating its genome, we aim to expand future evolutionary studies on the *Caenorhabditis* genus.



Property	Metric
Location found	Da Lat, Vietnam
Location coordinates	12.17523, 108.69884

Mating system	Dioecious
Genome size (Mb)	126
Approximate coverage (x) – PacBio HiFi	171
Approximate coverage (x) – Illumina Hi-C	643
Number of reads (Gb) – PacBio HiFi	21.6
Number of reads (Gb) – Illumina Hi-C	81.3
GC content (%)	37.96
Number of protein coding genes	18,221
Repeats (%) – RepeatMasker	20.11
Repeats (%) – EarlGrey	24.6
SRA accession no. – PacBio HiFi	SRR38757051
SRA accession no. – Illumina Hi-C	SRR38757052
Isolated by	Marie-Anne Félix lab

Figure 1. PacBio and Hi-C sequencing leads to a chromosome-scale assembly of novel *Caenorhabditis* sp. 65:

A. Snail plot (BlobTk) of assembly statistics for JU4118. The outer circumference of the plot represents the full length of the genome. The rings, from outer- to inner-most describe the features: the blue rings represent %GC content per scaffold; the pale orange ring represents N90 length in Mb; the bold orange ring represents N50 length in Mb; the red wedge is the longest scaffold length in Mb; the dark gray represents the length of each scaffold, arranged from longest to shortest in a clockwise direction; the light gray represents total scaffold count on a logarithmic scale. The top right circle describes BUSCO scores, displaying a summary of complete, duplicated, fragmented, and missing genes in the nematoda_odb12 set. **B.** Hi-C contact map of JU4118 assembly reveals the six chromosome-scale scaffolds (X sex chromosome, autosomes I-V, ordered from largest to smallest. Scaffolds 7-10 are unplaced **C.** Synteny plot of single-copy orthologs from pairwise comparisons of *C. sp. 65*, *C. elegans*, and *C. doughertyi*. Colors correspond to *C. elegans* chromosomes. **D.** *Caenorhabditis* phylogeny of Elegans Group, with *C. japonica* as the outgroup species. **E.** Percent of nucleotides annotated as genes (red), repeats (blue), or GC content (green) in 10 kb bins across each chromosome. **F.** Chromosomes with 100 kb binned counts of single-copy orthologs to *C. elegans* genes previously assigned to Nigon elements (A-E, N, X). **G.** Kimura-2 parameter distance showing TE "age" of the major TE subclasses annotated in JU4118.

Table 1. Properties of *Caenorhabditis* sp. 65

Description

Caenorhabditis is a genus of the phylum of Nematoda that consists of diverse roundworms, including the species *Caenorhabditis elegans*. *C. elegans* has become an important model organism in biology. In an effort to understand further evolutionary biology of *Caenorhabditis*, many species have been sampled around the world to be sequenced and described (Stevens et al., 2019). Here, we aim to describe *Caenorhabditis* sp. 65 (strain JU4118). Morphologically, *C. sp. 65* is similar in size and appearance to *C. elegans* when viewed under a stereo-microscope. The mating systems of the two species, though, differ. *C. sp. 65* is a dioecious species, whereas *C. elegans* is androdioecious. *C. sp. 65* was isolated from rotting fruit in Da Lat, Vietnam by the Marie-Anne Félix lab (Table 1) and was inbred for 25 generations. Phylogenetic analysis places *C. sp. 65* as a sister to *C. doughertyi*, an outcrossing species with a genome size of ~147 Mb (Figure 1, Panel D).

Using PacBio HiFi long-read sequencing (~171× coverage) and Hi-C Illumina paired-end sequencing (~643× coverage), assembly with Hifiasm (Cheng et al., 2021) resulted in 471 contigs spanning ~203 Mb. After removal of bacterial contigs using BLAST (Camacho et al., 2009), purging haplotigs with purge_dups (Guan et al., 2020), and scaffolding with YaHS (Zhou et al., 2023), the final genome assembly was 126 Mb in 10 scaffolds (Figure 1, Panel A). Scaffolds were assigned chromosomal identities based on synteny with *C. elegans*, revealing five autosomes, the X chromosome, and four smaller unplaced scaffolds (Figure 1, Panel B). Nucleotide BLAST (Camacho et al., 2009) reveals that three of the four unplaced scaffolds have similarity to several known *Caenorhabditis* species. The other unplaced scaffold contains about 98% repeats and does not have similarity to any known sequences.

The chromosome-level assembly of *C. sp. 65* is highly syntenic with *C. elegans* and shows similar patterns of chromosomal organization, consistent with other complete genomes in the Elegans Group. Identification of Nigon elements, ancestral linkage groups that show macro-syntenic changes resulting in present-day karyotypes, reveals the

same pattern as that seen across *Caenorhabditis*: nigons A, B, C, D, and E correspond to chromosomes I, II, III, IV, and V, respectively, while the X is a fusion between nigons N and X (Figure 1, Panel E) (Tandonnet et al., 2019). Across all chromosomes, repeat sequences generally appear higher in density on the first and last quarter of the chromosome, often referred to as the chromosome “arms.” Appearance of genes displays the opposite pattern: genes are more frequent in chromosome centers (Figure 1, Panel F). GC content also has a similar landscape across the chromosomes as compared to *C. elegans*, hovering around 38% across the genome (Correa et al., 2025).

Despite the conservation of macrosynteny, within chromosomes there are several noticeable rearrangements, inversions, and expansions (Figure 1, Panel C). For example, *C. sp. 65* shows an expansion on the X chromosome relative to *C. elegans*. This enlarged X chromosome, and the enlarged genome in general, seems to be a feature of this clade as it is also present in *C. doughertyi* and *C. sp. 61* (Lad et al., 2026). The genome sizes of *C. elegans*, *C. sp. 65*, *C. sp. 61*, and *C. doughertyi* are 100 Mb, 126 Mb, 165 Mb, and 147 Mb, respectively. The X chromosome lengths are 18 Mb, 30 Mb, 42 Mb, and 32 Mb, respectively. Dioecious species are known to have larger genome sizes than androdioecious species, and several theories exist as to why this might be, including: gene family turnover (Adams et al., 2023), accumulation of repetitive elements (Woodruff and Teterina, 2020), and intron size variation (Stevens et al., 2019).

The *C. sp. 65* predicted gene set had a BUSCO (Manni et al., 2021) completeness of 99.5%. The assembly contained 18,221 predicted genes, making up 31.5% of the genome. This is slightly less than *C. elegans*, which has approximately 20,000 genes. Accordingly, there were more *C. elegans*-specific and *C. sp. 65*-reduced orthogroups than *C. sp. 65*-specific and *C. sp. 65*-expanded orthogroups. GO term analysis of OrthoFinder (Emms & Kelly, 2019) output shows that *C. sp. 65*-specific genes are enriched in processes underlying embryo development, negative regulation of vulval development, innate immune response, and ubiquitin-dependent protein degradation. *C. elegans*-specific genes are enriched in processes related to sensing the environment, including GPCR signaling pathways, sensory perception of a chemical stimulus, and olfactory behavior. However, many genes were lacking in functional annotation. For example, the region of expansion on the X chromosome showed many genes with no functional domains annotated, and those that were, tended to be predicted disordered regions.

The *C. sp. 65* assembly was annotated by EarlGrey (Baril et al., 2024) to be composed of 24.6 % repetitive elements, slightly more than *C. elegans* 20.5 %. Percent repeats for each chromosome of *C. sp. 65* was: 25.5 %, 19.2 %, 24.9 %, 22.5 %, 18.8 %, and 32.7 %, in order of chromosome I, II, III, IV, V, and X. Similarly, the percent repeats across *C. elegans* was: 23.4 %, 21.2 %, 24.8 %, 19.1 %, 21.6 %, and 14.1 %. The X chromosome of *C. sp. 65* had a percentage of repeats twice that of *C. elegans*. The same pattern has been seen in *C. sp. 61* (Lad et al., 2026). Repeat classifications showed more long interspersed nuclear elements (LINEs) (1.3 % vs 0.6 %), long terminal repeats (LTRs) (1.0 % vs 0.4 %), simple repeats (8.0 % vs 4.6 %), and unclassified repeats (10.6 % vs 4.0 %) in *C. sp. 65* compared to *C. elegans*. Kimura distance (Kimura, 1980) which is often used as a proxy for transposable element “age” but simply a measure of sequence divergence, reveals periodic small bursts of LINEs, and a recent expansion of LINEs, LTRs, DNA transposons, and unclassified repeats (Figure 1, Panel G).

Telomeric-repeat motifs (TRMs) can also be found in this assembly. Telomeres are composed of tandem arrays of TRMs that are recognized by telomere-binding proteins that ensure replication and protection of chromosomal ends. TRMs are conserved in Nematoda, with the canonical TRM sequence of TTAGGC (Lim et al., 2023). We found that all chromosomes, except I and V, in the JU4118 assembly contain the canonical TRM at the 5' terminus. At the 3' terminus, all chromosomes, except II, contain the canonical sequence. The unplaced scaffolds do not appear to have TRM sequences at either terminus.

More work is needed to understand the evolutionary forces shaping *Caenorhabditis* genomes, both within this clade and across the genus in general. Our draft genome of *C. sp. 65*, JU4118 provides a resource for these future molecular, comparative, and phylogenetic evolutionary studies.

Methods

Collection: Worms were kept in cryopreservation until transfer to our research group and were maintained continuously thereafter using standard *Caenorhabditis elegans* techniques. Worms were grown on agar plates at 20°C using nematode growth media and seeded with a lawn of *Escherichia coli* strain OP50. For DNA extractions, worm populations were expanded by transferring a small “chunk” of agar to three 100 mm plates seeded with *E. coli* and left at 20°C for 2-3 days until plate was filled with mixed-age worms but free of dauer larvae. Worms were then washed off plates with M9 buffer into a 15 mL conical tube and washed twice with M9 buffer to minimize surface contaminants, then resuspended in 10 mL M9 and left on a rocker overnight (~17 hours) to purge gut of further contaminants. Prior to extraction, worms were washed twice more with M9. Worms were collected by pelleting worm bodies via centrifugation, removing M9 supernatant, and transferring 50 µL aliquots of worm pellet to 1.5 mL tubes.

Long-read sequencing: DNA for long-read sequencing was extracted using the Promega Wizard® HMW DNA Extraction Kit (cat. no. A2920), using the manufacturer protocol with minor modifications. Worm cuticles were broken by repeated freeze/thaw cycles where one tube of live worm pellet was placed at -80°C for five minutes, moved to 37°C until

10/4/2026 - Open Access

thawed, briefly vortexed, and then frozen again, for five cycles. All centrifuge steps were done at 4°C and alcohols were kept on ice until use. At the lysis step, an extra incubation step of 25 minutes at 65°C was added. PacBio sequencing was performed at the University of Miami's John P. Hussman Institute for Human Genomics Sequencing Core Facility (RRID:SCR_017828).

Hi-C: Extra tubes of 50 µL worm pellet were frozen at -80°C using a Mr. Frosty™ freezing container (Thermo Scientific cat. no. 5100-001) filled with 100% isopropanol to prevent ice crystal formation. Two tubes were then mailed on dry ice to Arima Genomics for High Coverage Chromatin Conformation Capture sequencing (Hi-C).

Genome Assembly: PacBio HiFi and Arima Hi-C libraries were assembled using Hifiasm v0.16.0 (Cheng et al., 2021) using default parameters. BLAST v2.14.1 (Camacho et al., 2009) was used to identify and remove contaminant contigs from the diploid and phased haploid assemblies. PacBio HiFi reads were mapped to the assembly using minimap2 v2.30 (Li, 2018), with parameter -x map-hifi and read depth cutoffs were calculated with pbcstat from purge_dups v1.2.6 (Guan et al., 2020). Assembly self-alignment was performed with minimap2 v2.30 and parameters -xasm5 -DP to find duplications. Alternative haplotypes were subsequently removed with purge_dups v1.2.6. To quality check the assemblies between each step, QUAST v5.3.0 (Gurevich et al., 2013) was used with default parameters and BUSCO v6.0.0 (Manni et al., 2021) was run against lineage dataset Nematoda odb12 with option -m genome and --offline.

Phylogenetic Analysis: A multigene tree was calculated from the BUSCO output of 95 species within Rhabditidae. Single copy orthologs were concatenated for multiple sequence alignment, which was completed with MAFFT v7.221 using the localpair option with 1000 times iterative refinement (Katoh and Standley, 2013). Alignments were trimmed using ClipKIT v2.3.0 with option smart-gap (Steenwyk et al., 2020) and partitioned with AMAS concat (Borowiec, 2016). IQ-TREE v1.6.12 was used to compute the phylogenetic tree with option -m MFP (Nguyen et al., 2015). Branch length supports were calculated with SH-like approximate likelihood ratio test and 1000 bootstrap replicates (Guindon et al., 2010). Visualization was done in R with package phytools (Revell, 2012). Tips containing species too distant from *C. sp.* 65 were dropped for clarity.

Hi-C Mapping: Juicer v2.0 (Durand et al., 2016) was used for alignment and processing of Hi-C raw data using default parameters, with the --assembly option. YaHS v1.2.2 (Zhou et al., 2023) was used for scaffolding. Juicebox v2.3.6 (Robinson et al., 2018) was used for visualization and assessment.

Gene and Repeat Annotation: Gene annotation was performed with BRAKER3 v3.0.8 (Gabriel et al., 2024) on the softmasked assemblies using the protein dataset Nematoda odb10 and RNA sequence data downloaded from NCBI project [PRJNA1256413](#) (O'Leary et al., 2024). Prior to BRAKER3, genomes were softmasked with RepeatModeler2 (Flynn et al., 2020) and RepeatMasker (Smit et al., 2013), and RNA reads were aligned to the genome with STAR v2.6.1a and option --outSAMstrandField intronMotif (Dobin et al., 2013). Briefly, BRAKER relies on 2 generalized hidden markov models for gene prediction, GeneMark (unsupervised) and AUGUSTUS (supervised) (Brůna et al., 2024; Stanke et al., 2006). The resulting protein sets are then combined by TSEBRA (Gabriel et al., 2021) to maximize BUSCO completeness scores. Protein predictions were then filtered for the longest isoform using AGAT v1.4.1 (Dainat et al., 2026), specifically the scripts `agat_sp_keep_longest_isoform.pl` and `agat_sp_extract_sequences.pl`. Statistics were generated with `agat_sp_statistics.pl`. Functional annotations were done with InterproScan v5.68.100.0 with options -dp -goterms -pathways (Jones et al., 2014). OrthoFinder v2.5.5 (Emms & Kelly, 2019) was used to find single copy orthologs between JU4118 and *C. elegans*. The six major chromosomes were identified by location of single copy orthologs on *C. elegans* chromosomes. Nigon element classifications were assigned to single copy orthologs using a list of known gene:Nigon associations from Gonzalez de la Rosa (2021). Finally, GO term enrichment analysis was conducted in R with package topGO (Alexa & Rahnenführer, 2026).

Repetitive elements were annotated with EarlGrey v6.0.1 (Baril et al., 2024) and options -r nematoda -e yes. Briefly, EarlGrey employs a BLAST, Extract, Align, Trim (BEAT) process adapted from TEStainer (<https://github.com/jamesdgalbraith/TEstainer>) along with subprocesses: Tandem Repeat Finder (Benson, 1999), MREPS (Kolpakov et al., 2003) SA-SSR (Pickett et al., 2013), LTR_FINDER (Xu & Wang, 2007), RepeatModeler2, and RepeatMasker. EarlGrey merges and defragments the annotations with RepeatCraft (Wong & Simakov, 2019) to produce the final consensus library.

Snail Plots: Snail plots were generated by BlobTk v0.8.0 (Challis & Blaxter, 2026) using the soft-masked assemblies. Assembly BUSCO scores were recalculated for the snail plots using BUSCO v6.0.0 (Manni et al., 2021) and the nematoda_odb12 dataset.

Data availability

Bioinformatic scripts, workflows and software commands are available at https://github.com/jannafierst/HiC_Assemblies

Scaffolded genome and annotation supplementary files are available at <https://zenodo.org/records/21998323>

Nucleotide sequence accession numbers

DNA libraries used in this project have been deposited at the Sequence Read Archive (SRA): SRR38757051 (PacBio) and SRR38757052 (Hi-C) under NCBI under Bioproject [PRJNA1256413](#).

Acknowledgements: We would like to give special thanks to the Marie-Anne Félix lab for collecting and providing samples.

References

Adams PE, Eggers VK, Millwood JD, Sutton JM, Pienaar J, Fierst JL. 2023. Genome Size Changes by Duplication, Divergence, and Insertion in *Caenorhabditis* Worms. *Molecular Biology and Evolution* 40: 10.1093/molbev/msad039. DOI: [10.1093/molbev/msad039](#)

Alexa A, Rahnenführer J (2026). *topGO: Enrichment Analysis for Gene Ontology*. doi:10.18129/B9.bioc.topGO. R package version 2.64.0, <https://bioconductor.org/packages/topGO>.

Baril T, Galbraith J, Hayward A. 2024. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. *Mol Biol Evol* 41(4): 10.1093/molbev/msae068. PubMed ID: [38577785](#)

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 27(2): 573-80. PubMed ID: [9862982](#)

Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4: e1660. PubMed ID: [26835189](#)

Brûna T, Lomsadze A, Borodovsky M. 2024. GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. *Genome Res* 34(5): 757-768. PubMed ID: [38866548](#)

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421. PubMed ID: [20003500](#)

Challis R, Blaxter M. 2026. Snail plots are badges of genome assembly quality. *G3: Genes, Genomes, Genetics* : 10.1093/g3journal/jkag074. DOI: [10.1093/g3journal/jkag074](#)

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* 18: 170-175. DOI: [10.1038/s41592-020-01056-5](#)

Correa JS, Noble LM, Sloat SA, Nguyen THM, Rockman MV. 2025. Conservative evolution of genetic and genomic features in *Caenorhabditis* *becei*, an experimentally tractable gonochoristic worm. *bioRxiv*: pii: 2025.05.09.653148. 10.1101/2025.05.09.653148. PubMed ID: [40463230](#)

Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format. Zenodo.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al., Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1): 15-21. PubMed ID: [23104886](#)

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* 3(1): 95-8. PubMed ID: [27467249](#)

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20(1): 238. PubMed ID: [31727128](#)

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* 117(17): 9451-9457. PubMed ID: [32300014](#)

Gabriel L, Brûna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2024. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res* 34(5): 769-777. PubMed ID: [38866550](#)

Gabriel L, Hoff KJ, Brûna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* 22(1): 566. PubMed ID: [34823473](#)

Gonzalez de la Rosa PM, Thomson M, Trivedi U, Tracey A, Tandonnet S, Blaxter M. 2021. A telomere-to-telomere assembly of *Oscheius tipulae* and the evolution of rhabditid nematode chromosomes. *G3 (Bethesda)* 11(1): 10.1093/g3journal/jkaa020. PubMed ID: [33561231](#)

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36(9): 2896-2898. PubMed ID: [31971576](#)

Guindon Sp, Dufayard JFo, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology* 59: 307-321. DOI: [10.1093/sysbio/syq010](#)

- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29: 1072-1075. DOI: [10.1093/bioinformatics/btt086](https://doi.org/10.1093/bioinformatics/btt086)
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al., Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30(9): 1236-40. PubMed ID: [24451626](https://pubmed.ncbi.nlm.nih.gov/24451626/)
- Kapitonov VV, Jurka J. 2008. A universal classification of eukaryotic transposable elements implemented in Repbase. *Nat Rev Genet* 9(5): 411-2; author reply 414. PubMed ID: [18421312](https://pubmed.ncbi.nlm.nih.gov/18421312/)
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30(4): 772-80. PubMed ID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16(2): 111-20. PubMed ID: [7463489](https://pubmed.ncbi.nlm.nih.gov/7463489/)
- Kolpakov R, Bana G, Kucherov G. 2003. mreps: Efficient and flexible detection of tandem repeats in DNA. *Nucleic Acids Res* 31(13): 3672-8. PubMed ID: [12824391](https://pubmed.ncbi.nlm.nih.gov/12824391/)
- Lad P, McCauley MA, Eggers VK, Fierst JL, Willcott K. 2026. Chromosome-Scale Assembly of Novel *Caenorhabditis* sp. 61 (strain JU4110). *MicroPubl Biol* 2026: 10.17912/micropub.biology.002358. PubMed ID: [42798900](https://pubmed.ncbi.nlm.nih.gov/42798900/)
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34: 3094-3100. DOI: [10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191)
- Lim J, Kim W, Kim J, Lee J. 2023. Telomeric repeat evolution in the phylum Nematoda revealed by high-quality genome assemblies and subtelomere structures. *Genome Res* 33(11): 1947-1957. PubMed ID: [37918961](https://pubmed.ncbi.nlm.nih.gov/37918961/)
- Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc* 1(12): e323. PubMed ID: [34936221](https://pubmed.ncbi.nlm.nih.gov/34936221/)
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2014. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution* 32: 268-274. DOI: [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300)
- O'Leary NA, Cox E, Holmes JB, Anderson WR, Falk R, Hem V, et al., Schneider VA. 2024. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. *Sci Data* 11(1): 732. PubMed ID: [38969627](https://pubmed.ncbi.nlm.nih.gov/38969627/)
- Pickett BD, Karlinsey SM, Penrod CE, Cormier MJ, Ebbert MT, Shiozawa DK, Whipple CJ, Ridge PG. 2016. SA-SSR: a suffix array-based algorithm for exhaustive and efficient SSR discovery in large genetic sequences. *Bioinformatics* 32(17): 2707-9. PubMed ID: [27170037](https://pubmed.ncbi.nlm.nih.gov/27170037/)
- Revell LJ. 2011. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3: 217-223. DOI: [10.1111/j.2041-210X.2011.00169.x](https://doi.org/10.1111/j.2041-210X.2011.00169.x)
- Robinson JT, Turner D, Durand NC, Thorvaldsdóttir H, Mesirov JP, Aiden EL. 2018. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. *Cell Syst* 6(2): 256-258.e1. PubMed ID: [29428417](https://pubmed.ncbi.nlm.nih.gov/29428417/)
- Smit A, Hubley R, Green P. 2013. RepeatMasker. <http://www.repeatmasker.org>
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34(Web Server issue): W435-9. PubMed ID: [16845043](https://pubmed.ncbi.nlm.nih.gov/16845043/)
- Steenwyk JL, Buida TJ, Li Y, Shen XX, Rokas A. 2020. ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference. *PLOS Biology* 18: e3001007. DOI: [10.1371/journal.pbio.3001007](https://doi.org/10.1371/journal.pbio.3001007)
- Stevens L, Félix MA, Beltran T, Braendle C, Caurcel C, Fausett S, et al., Blaxter. 2019. Comparative genomics of 10 new *Caenorhabditis* species. *Evolution Letters* 3: 217-236. DOI: [10.1002/evl3.110](https://doi.org/10.1002/evl3.110)
- Tandonnet S, Koutsovoulos GD, Adams S, Cloarec D, Parihar M, Blaxter ML, Pires-daSilva A. 2019. Chromosome-Wide Evolution and Sex Determination in the Three-Sexed Nematode *Auanema rhodensis*. *G3 Genes|Genomes|Genetics* 9: 1211-1230. DOI: [10.1534/g3.119.0011](https://doi.org/10.1534/g3.119.0011)
- Wong WY, Simakov O. 2019. RepeatCraft: a meta-pipeline for repetitive element de-fragmentation and annotation. *Bioinformatics* 35(6): 1051-1052. PubMed ID: [30165587](https://pubmed.ncbi.nlm.nih.gov/30165587/)
- Woodruff GC, Teterina AA. 2020. Degradation of the Repetitive Genomic Landscape in a Close Relative of *Caenorhabditis elegans*. *Mol Biol Evol* 37(9): 2549-2567. PubMed ID: [32359146](https://pubmed.ncbi.nlm.nih.gov/32359146/)
- Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35(Web Server issue): W265-8. PubMed ID: [17485477](https://pubmed.ncbi.nlm.nih.gov/17485477/)
- Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* 39(1): 10.1093/bioinformatics/btac808. PubMed ID: [36525368](https://pubmed.ncbi.nlm.nih.gov/36525368/)

10/4/2026 - Open Access

Funding:

Supported by National Institute of General Medical Sciences (United States) GM147245 to Janna L. Fierst.

Conflicts of Interest: The authors declare that there are no conflicts of interest present.

Author Contributions: Jasbelle Sosa: writing - original draft, writing - review editing. Michelle A. McCauley: data curation, formal analysis, visualization, writing - review editing. Victoria K. Eggers: formal analysis, visualization, writing - review editing. Janna L. Fierst: conceptualization, formal analysis, funding acquisition, writing - review editing. Karolina Willicott: formal analysis, visualization, writing - review editing, project administration.

Reviewed By: Jose Salome Correa

Nomenclature Validated By: Anonymous

WormBase Paper ID: WBPaper00070214

History: Received May 22, 2026 **Revision Received** August 27, 2026 **Accepted** September 29, 2026 **Published Online** October 4, 2026 **Indexed** October 18, 2026

Copyright: © 2026 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Sosa J, McCauley MA, Eggers VK, Fierst JL, Willicott K. 2026. Chromosome-Scale Assembly of Novel *Caenorhabditis* sp. 65 (strain JU4118). microPublication Biology. [10.17912/micropub.biology.002448](https://doi.org/10.17912/micropub.biology.002448)