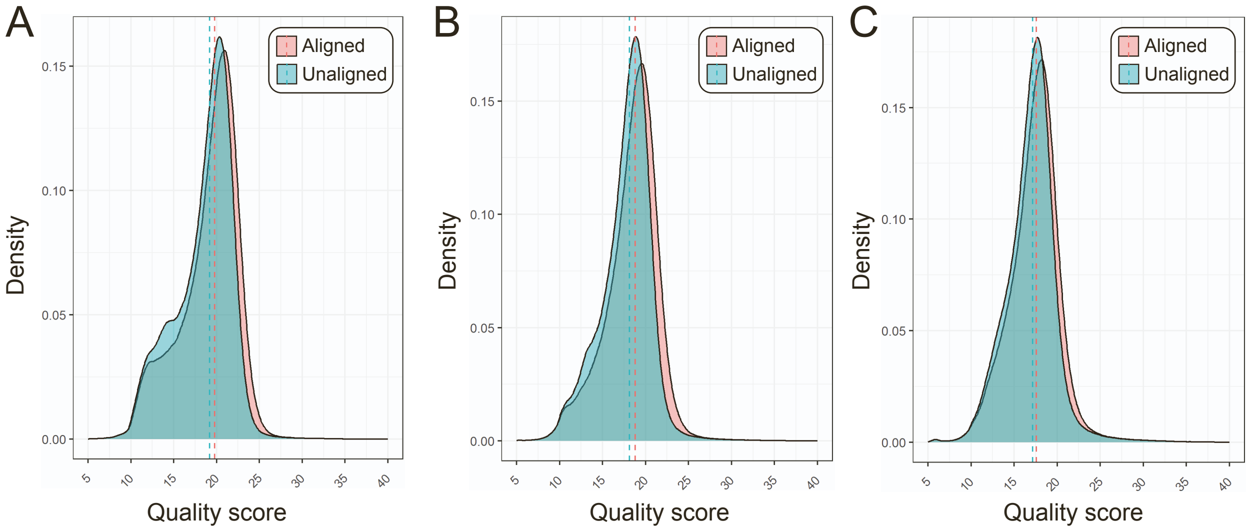
iScience, Volume *27*

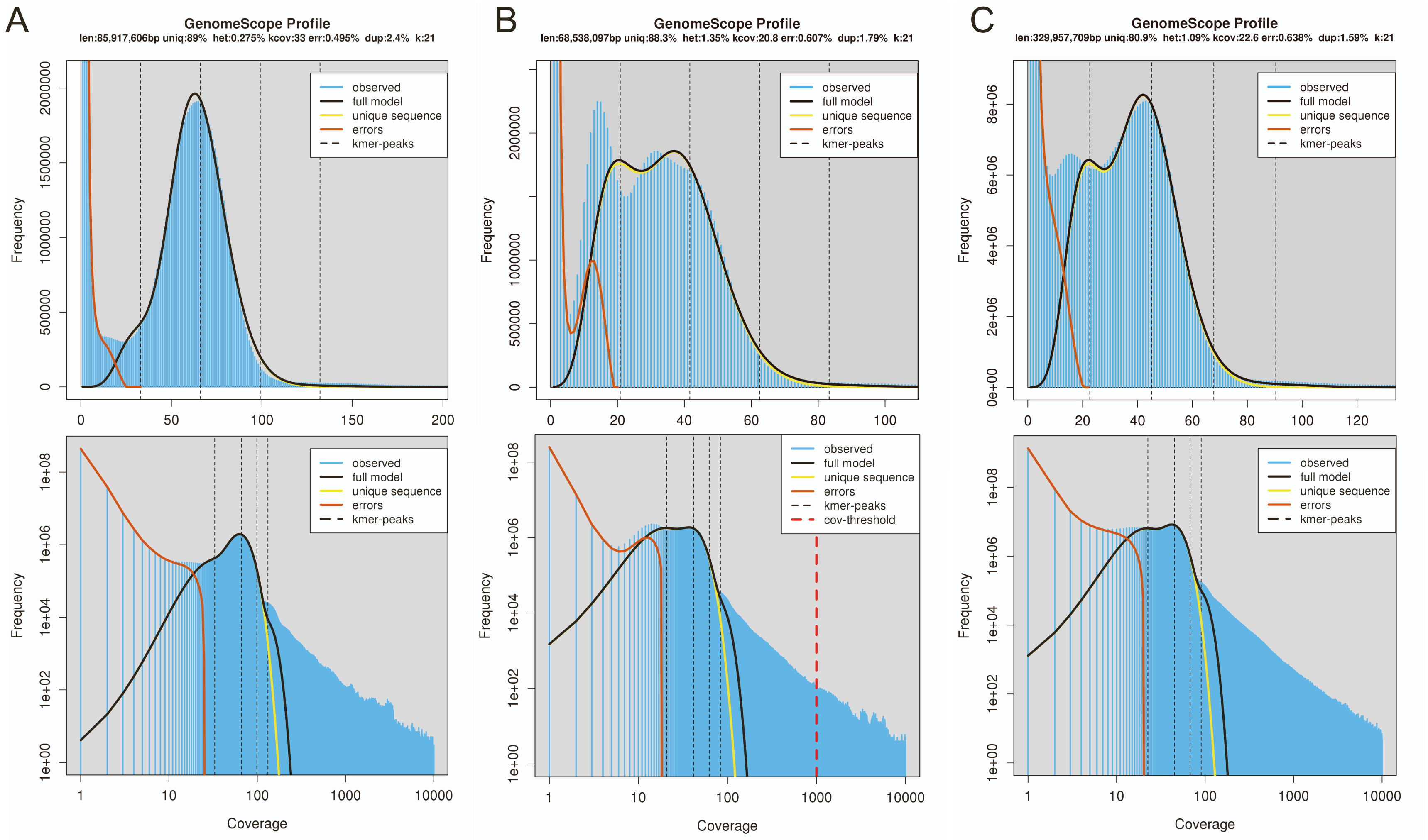
Supplemental information

Assessing *de novo* parasite genomes assembled using only Oxford Nanopore Technologies MinION data

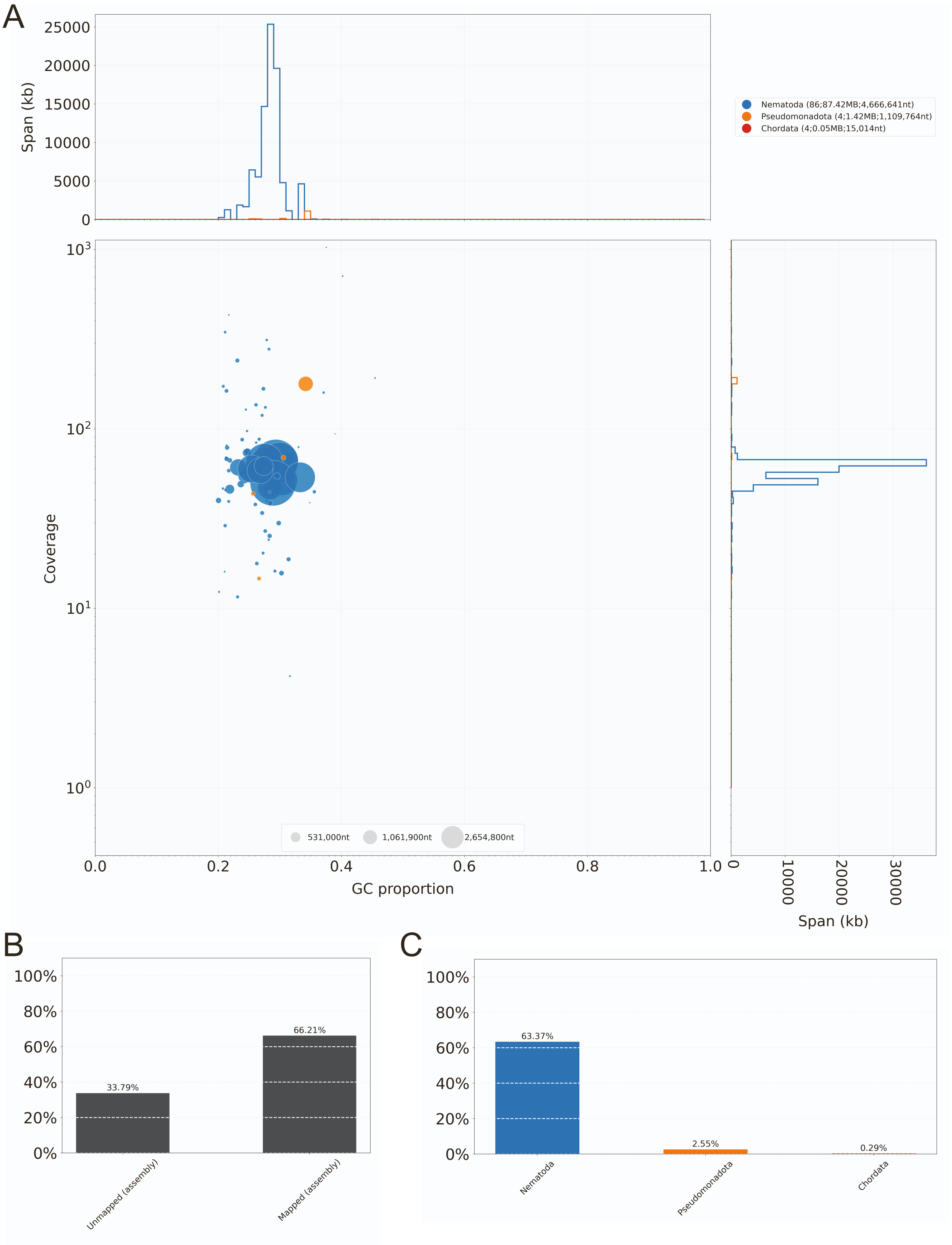
Kaylee S. Herzog, Rachel Wu, John M. Hawdon, Peter Nejsum, and Joseph R. Fauver



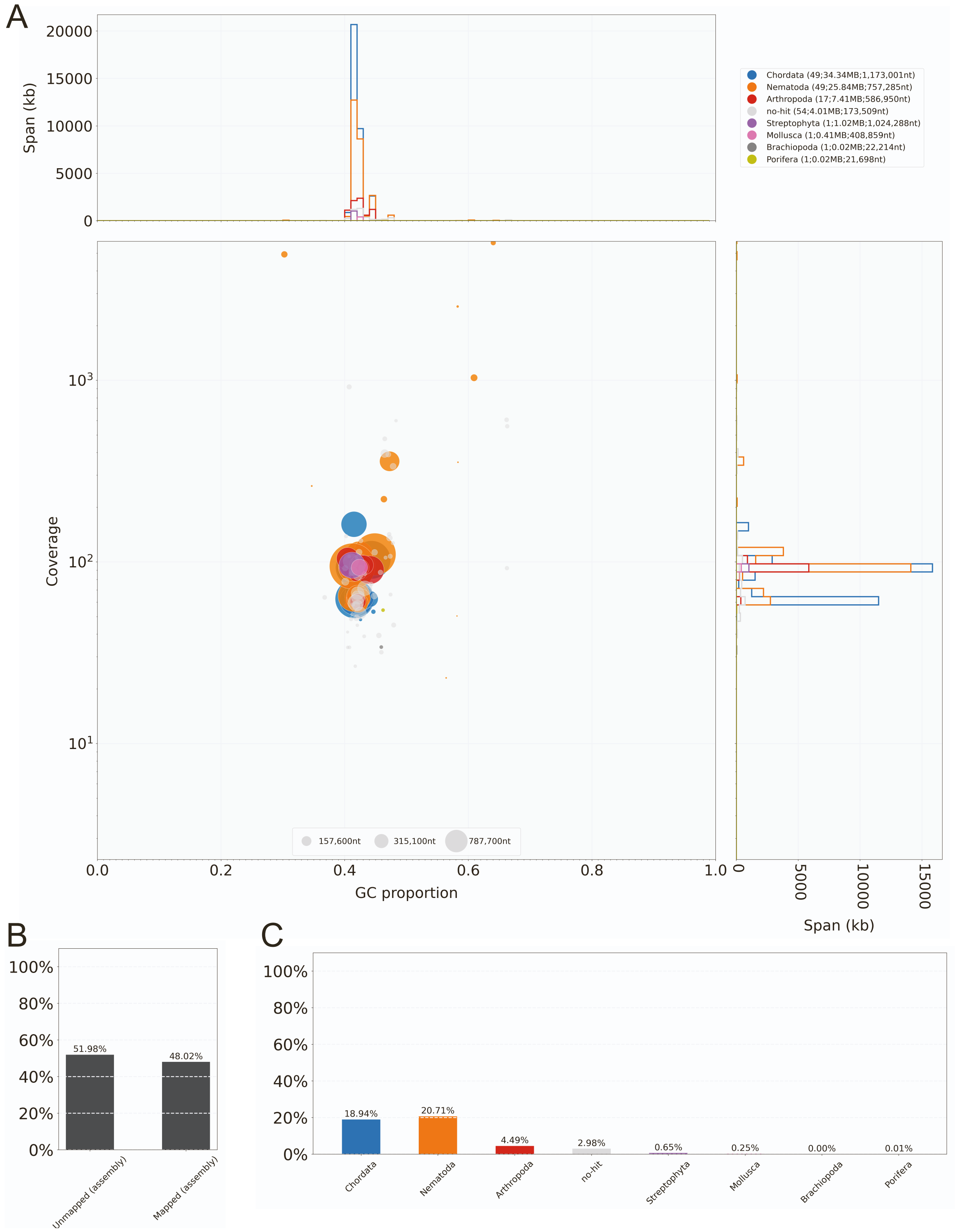
Supplemental Figure 1. Histograms of average read quality scores (i.e., Q scores) for unaligned basecalled MinION read data, and for basecalled MinION read data aligned to the final MinION data only genome assembly, related to STAR Methods. (A) *Brugia malayi*. (B) *Trichuris trichiura*. (C) *Ancylostoma caninum*.



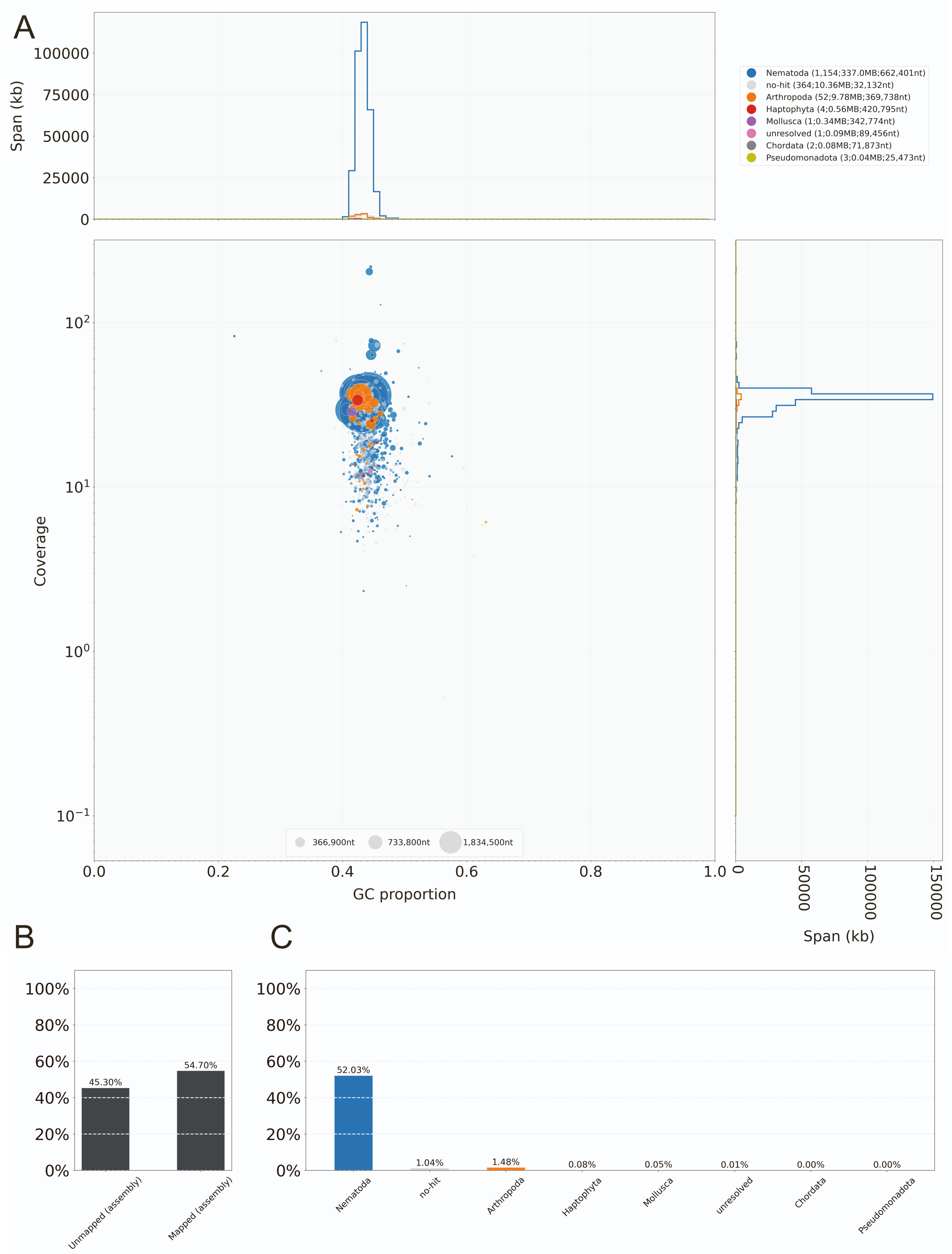
Supplemental Figure 2. **Focused (top) and complete (bottom) GenomeScope histograms based on quality-controlled Illumina read datasets generated for each species, realted to STAR Methods.** (A) *Brugia malayi*. (B) *Trichuris trichiura*. (C) *Ancylostoma caninum*.



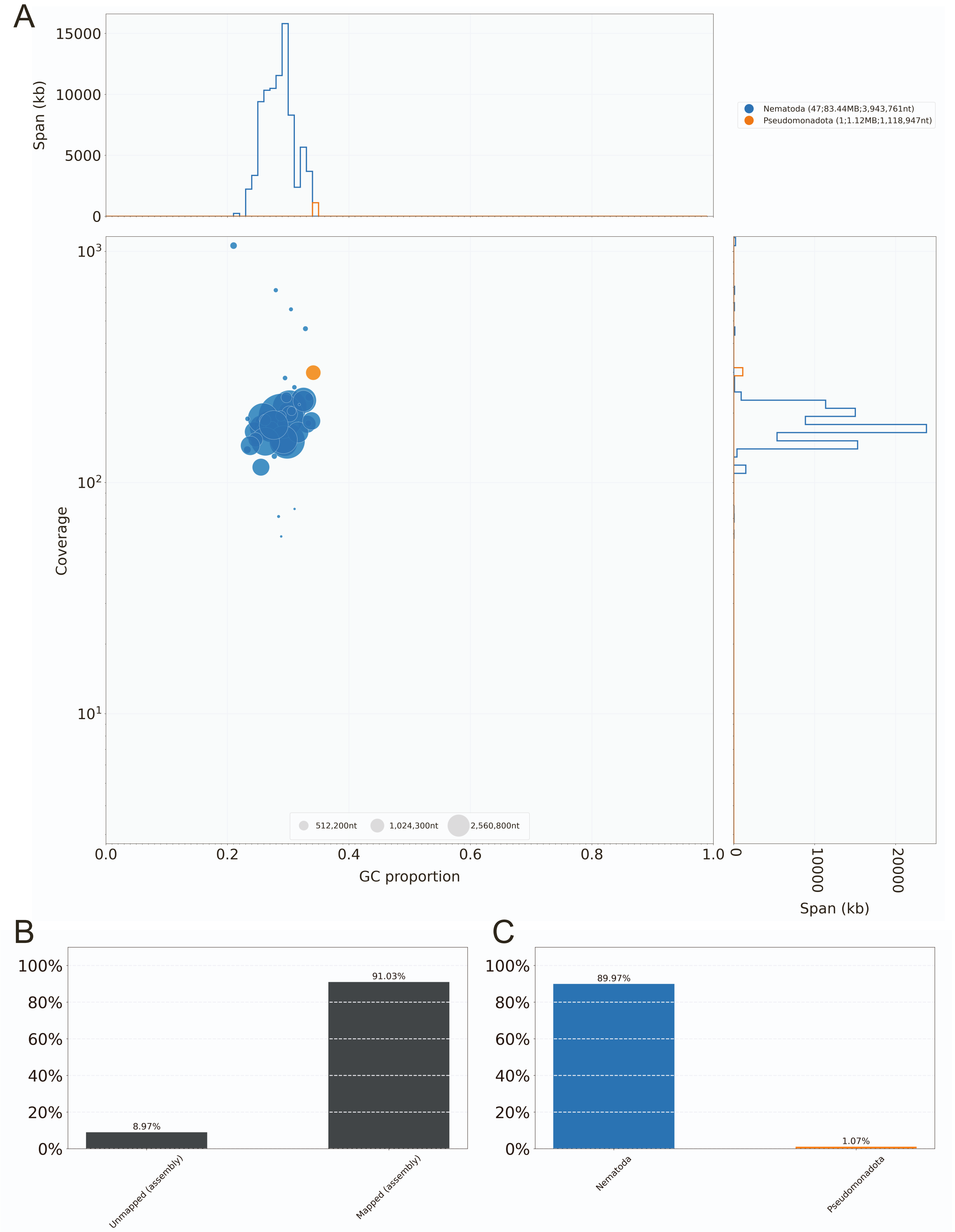
Supplemental Figure 3. BlobTools output for the MinION data-only assembly generated for *Brugia malayi*, related to STAR Methods. (A) BlobPlot. (B–C) Read coverage plots.



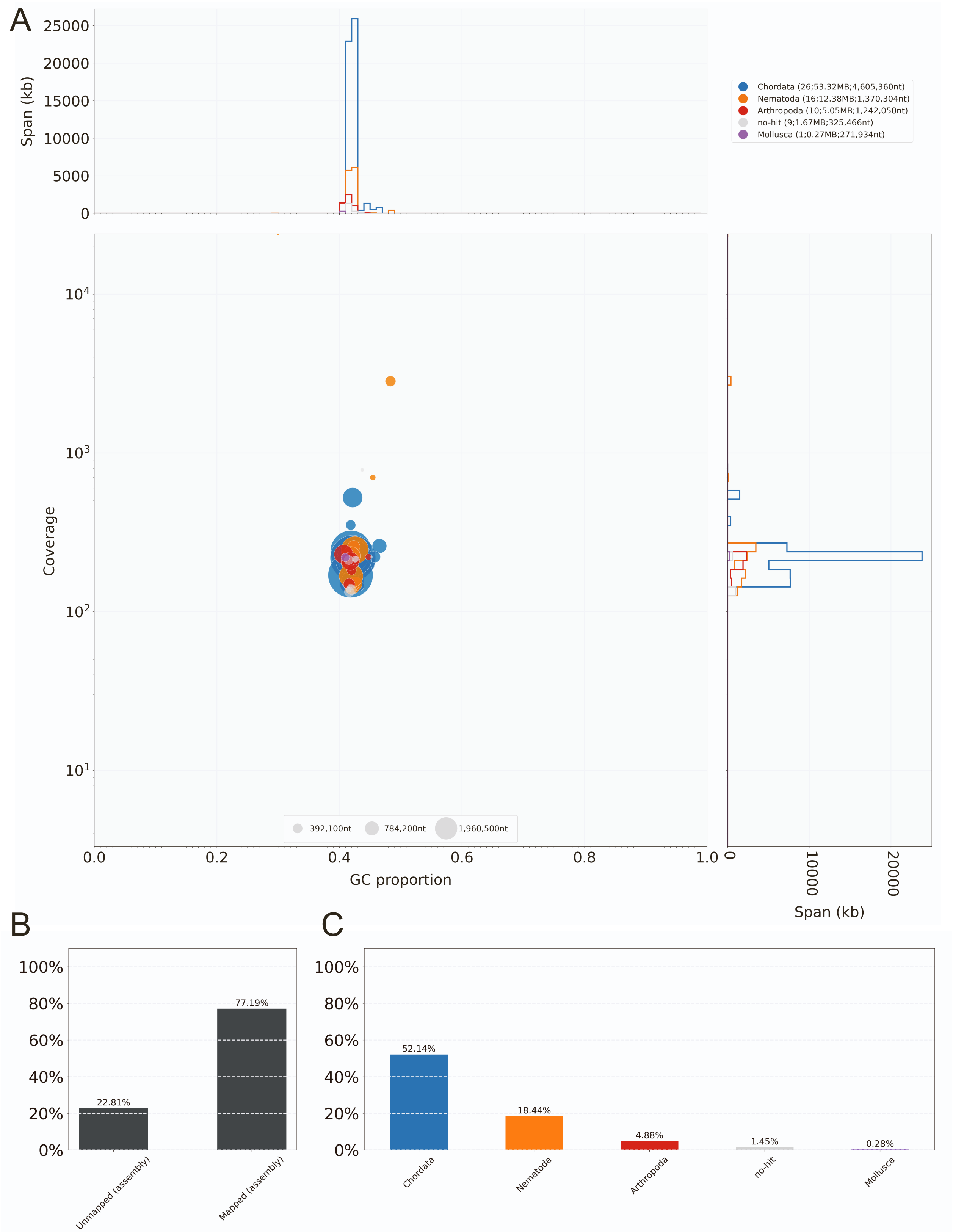
Supplemental Figure 4. BlobTools output for the MinION data-only assembly generated for *Trichuris trichiura*, realted to STAR Methods. (A) BlobPlot. (B–C) Read coverage plots.



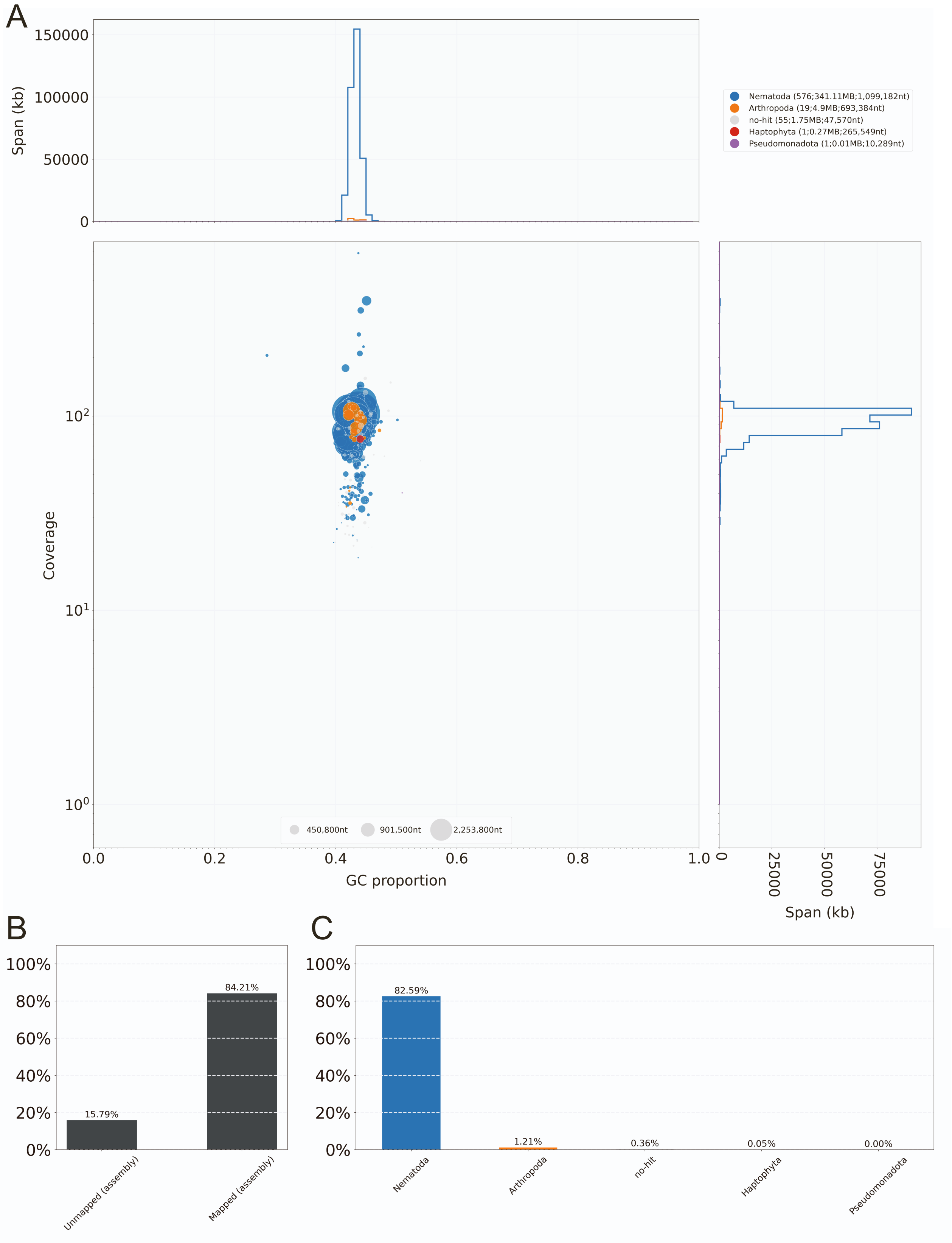
Supplemental Figure 5. BlobTools output for the MinION data-only assembly generated for *Ancylostoma caninum*, realted to STAR Methods. (A) BlobPlot. (B–C) Read coverage



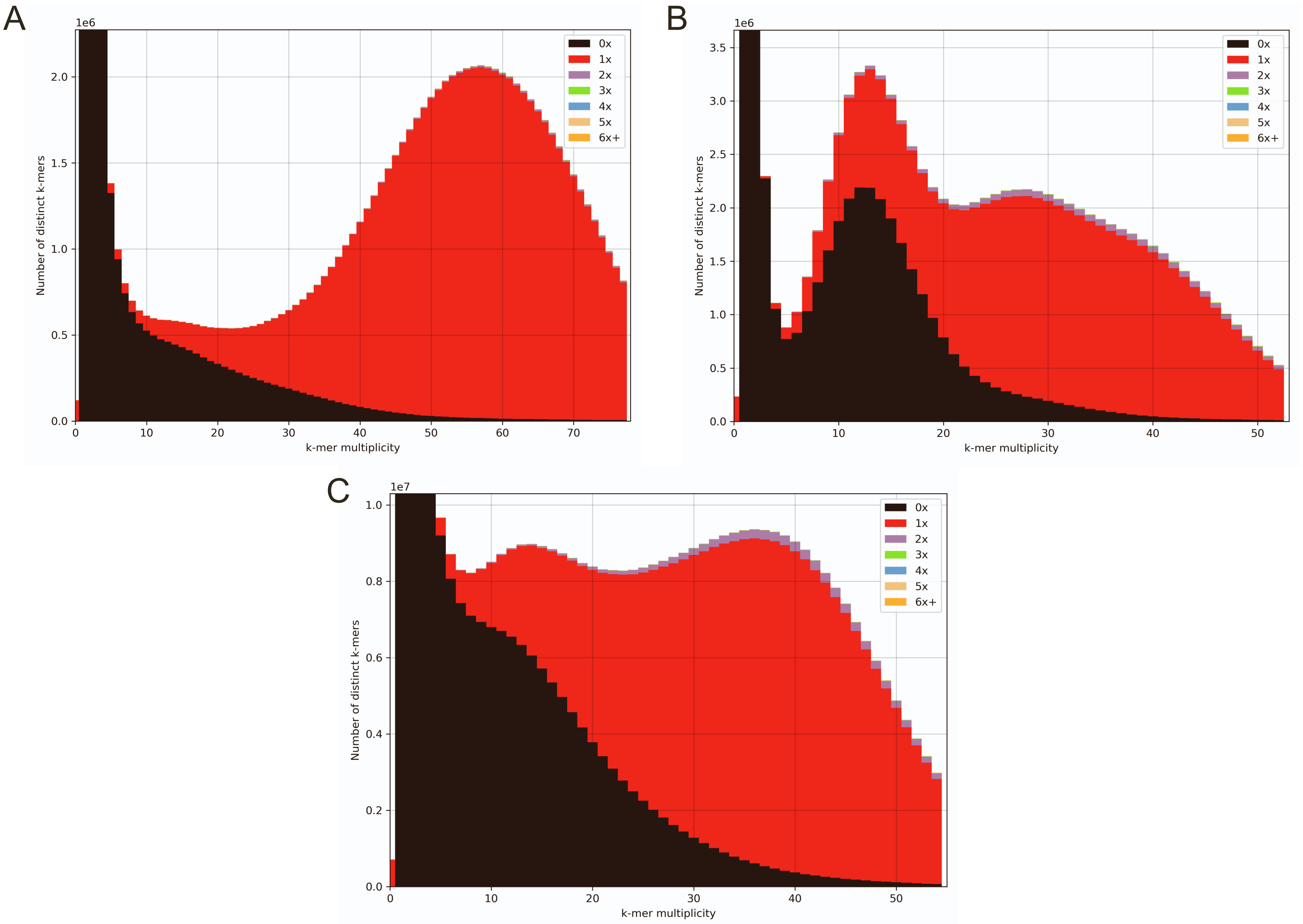
Supplemental Figure 6. BlobTools output for the hybrid assembly generated for *Brugia malayi*, related to STAR Methods. (A) BlobPlot. (B–C) Read coverage plots.



Supplemental Figure 7. BlobTools output for the hybrid assembly generated for *Trichuris trichiura*, realted to STAR Methods. (A) BlobPlot. (B–C) Read coverage plots.



Supplemental Figure 8. BlobTools output for the hybrid assembly generated for *Ancylostoma caninum*, realted to STAR Methods. (A) BlobPlot. (B–C) Read coverage plots.



Supplemental Figure 9. K-mer multiplicity plots generated from purged hybrid assemblies, realted to STAR Methods. (A) *Brugia malayi*. (B) *Trichuris trichiura*. (C) *Anyclostoma caninum*.

Supplemental Table 1. Extraction and ONT MinION and Illumina sequence data generation details for each of the focal species sequenced, related to STAR Methods. Asterisk (*) indicates amount of ONT MinION library remaining after the aliquot of the original library remaining following the first round of sequencing was rewashed with Long Fragment Buffer (LFB).

		Brugia malayi	***Trichuris trichiura***		***Ancylostoma caninum***		
Extraction	**Extracted material**	Single adult female	Single adult male		Pooled L3 larvae		
	Total amount of gDNA extracted (ng)	737	1,508		5,200		
	Mean fragment length of gDNA (bp)	>60,000	—		33,504		
Illumina data	**gDNA input for library prep (ng)**	42	40		400		
	No. PCR cycles	7	7		5		
	Data pre-quality control and filtering (bp)	7,992,851,592	4,567,580,276		20,844,376,730		
	Data post-quality control and filtering (bp)	7,850,596,985	4,442,501,155		20,179,886,997		
ONT MinION data	**Library name**	**Bm ♀ C**	**Tt ♂ 2D (all)**	**Tt ♂ 2D (LFB washed aliquot)**	**Acan L3 B1**	**Acan L3 B2**	**Acan L3 B1 + Acan L3 B2**
	Mean fragment length of gDNA post-additional bead cleanup (bp)	—	—	—	31,955	34,563	—
	Amount of gDNA input for library prep (ng)	500	1,259	—	406	399	—
	Amount of library generated (ng)	287	1,012	483*	189	264	—
	Amount of library sequenced (ng)	134	78.7	76.2	164	125	25 (B1) + 123 (B2)
	No. pores available at start of sequencing	1,413	1,544	1,635	1,527	1,465	1,538
	Total sequencing run time (hr)	72	72	62.75	80	67.75	80
	MinKNOW estimated read N50 (kb)	8.65	2.72	5.59	6.68	6.54	6.64
	MinKNOW estimated data generated (Gb)	15.91	14.06	12.76	10.86	10.73	11.56
	Data post-basecalling with Guppy (bp)	11,349,419,698	11,891,465,167	10,868,016,523	8,521,058,505	8,261,315,652	7,784,337,997
Assembly	**Est. depth of coverage for Illumina reads**	82.95×	48.58×		38.64×		
	Proportion of Illumina reads that mapped to reference assembly	95.74%	90.71%		90.31%		
	Est. depth of coverage for MinION reads	124.85×	249.91×		49.42×		
	Proportion of MinION reads that mapped to reference assembly	98.77%	99.91%		99.85%		
	GenomeScope estimated genome size (bp)	85,917,606	68,538,097		329,957,709		
	GenomeScope estimated heterozygosity	0.28%	1.35%		1.09%		

Supplemental Table 2. Comparison of different long read and hybrid *de novo* assemblers for *Brugia malayi*, related to STAR Methods. All QUAST and compleasm comparative data were generated from raw assemblies output by each assembler (i.e., from assemblies that were not purged for duplication, not polished, did not have contigs suspected as contamination removed, and did not have organellar genomes refined or added). “Popped” for the canu assembly indicates the raw assembly output by canu from which contigs indicated as potential alternative alleles (i.e., with FASTA headers including “suggestBubble=yes”) were removed.

		QUAST				Compleasm (Nematoda)			
		Size (bp)	**N50 (bp)**	**No. contigs**	**GC**	**Single copy**	**Duplicated**	**Fragmented**	**Missing**
Ghedin et al. (2007) reference		88,235,797	14,214,749	197	28.49%	3,097 (98.91%)	16 (0.51%)	10 (0.32%)	8 (0.26%)
MinION data only	**canu**	101,963,718	4,682,872	319	28.36%	2,979 (95.15%)	134 (4.28%)	10 (0.32%)	8 (0.26%)
	canu (popped)	95,068,481	6,405,829	195	28.47%	3,044 (97.22%)	69 (2.20%)	10 (0.32%)	8 (0.26%)
	wtdbg2	94,153,502	2,614,318	650	28.16%	2,930 (93.58%)	32 (1.02%)	13 (0.42%)	156 (4.98%)
	Flye	84,647,476	3,811,129	167	28.62%	3,097 (98.91%)	12 (0.38%)	11 (0.35%)	11 (0.35%)
	shasta	49,567,258	26,897	2,171	28.00%	1,515 (48.39%)	78 (2.49%)	81 (2.59%)	1,457 (46.53%)
Hybrid	**MaSuRCA**	86,029,419	2,961,229	57	28.56%	3,062 (97.80%)	50 (1.60%)	11 (0.35%)	8 (0.26%)
	WENGAN	81,095,327	2,941,126	105	28.59%	3,094 (98.82%)	8 (0.26%)	13 (0.42%)	16 (0.51%)
	HASLR	81,702,517	1,358,706	180	28.46%	3,026 (96.65%)	8 (0.26%)	12 (0.38%)	85 (2.71%)

Supplemental Table 3. Scores from compleasm for the assemblies generated as part of this study and the reference assemblies available for each species using the Metazoa and Eukaryota BUSCO databases, related to Table 2, Figure 2, and STAR Methods. Scores withing each category are presented as number of BUSCOs recovered in each assembly followed in parentheses by proportion of the total number of orthologs assessed by compleasm (954 and 255 for the Metazoa and Eukaryota databases, respectively).

		Metazoa Database				Eukaryota Database			
		Single	**Duplicated**	**Fragmented**	**Missing**	**Single**	**Duplicated**	**Fragmented**	**Missing**
Brugia malayi	**Ghedin et al. (2007) reference assembly**	772 (80.92%)	6 (0.63%)	13 (1.36%)	163 (17.09%)	251 (98.43%)	0 (0%)	1 (0.39%)	3 (1.18%)
	MinION data only assembly	775 (81.24%)	4 (0.42%)	12 (1.26%)	163 (17.09%)	251 (98.43%)	0 (0%)	1 (0.39%)	3 (1.18%)
	MinION assembly polished with Illumina data	775 (81.24%)	4 (0.42%)	12 (1.26%)	163 (17.09%)	251 (98.43%)	0 (0%)	1 (0.39%)	3 (1.18%)
	Hybrid assembly	775 (81.24%)	3 (0.31%)	12 (1.26%)	164 (17.19%)	250 (98.04%)	1 (0.39%)	1 (0.39%)	3 (1.18%)
Trichuris trichiura	**Foth et al. (2014) reference assembly**	667 (69.92%)	24 (2.52%)	23 (2.41%)	240 (25.16%)	217 (85.10%)	10 (3.92%)	8 (3.14%)	20 (7.84%)
	Doyle et al. (2022) reference assembly	667 (69.92%)	24 (2.52%)	23 (2.41%)	240 (25.16%)	217 (85.10%)	10 (3.92%)	8 (3.14%)	20 (7.84%)
	MinION data only assembly	676 (70.86%)	10 (1.05%)	22 (2.31%)	246 (25.79%)	219 (85.88%)	6 (2.35%)	8 (3.14%)	22 (8.63%)
	MinION assembly polished with Illumina data	676 (70.86%)	10 (1.05%)	22 (2.31%)	246 (25.79%)	218 (85.49%)	6 (2.35%)	9 (3.53%)	22 (8.63%)
	Hybrid assembly	679 (71.17%)	10 (1.05%)	21 (2.20%)	244 (25.58%)	222 (87.06%)	5 (1.96%)	7 (2.75%)	21 (8.24%)
Ancylostoma caninum	**International Helminth Genomes Consortium (2019) reference assembly**	643 (67.40%)	68 (7.13%)	37 (3.88%)	206 (21.59%)	215 (84.31%)	18 (7.06%)	14 (5.49%)	8 (3.14%)
	MinION data only assembly	678 (71.07%)	41 (4.30%)	19 (1.99%)	216 (22.64%)	228 (89.41%)	11 (4.31%)	6 (2.35%)	10 (3.92%)
	MinION assembly polished with Illumina data	681 (71.38%)	40 (4.19%)	20 (2.10%)	213 (22.33%)	228 (89.41%)	11 (4.31%)	6 (2.35%)	10 (3.92%)
	Hybrid assembly	693 (72.64%)	30 (3.14%)	21 (2.20%)	210 (22.01%)	234 (91.76%)	7 (2.75%)	6 (2.35%)	8 (3.14%)

Supplemental Table 4. Distance matrix for assemblies of the genome of the *Wolbachia* endosymbiont of *Brugia malayi* from both the MinION data only and hybrid assemblies generated herein compared to existing reference assemblies, related to STAR Methods. *Wolbachia* reference assemblies were obtained from Foster et al. (2005) (AE017321.1) and Lefoulon et al. (2019) (CP034333.1).

	AE017321.1 reference	**MinION data only assembly**	**CP034333.1 reference**	**Hybrid assembly**
AE017321.1 reference	——	23	39	184
MinION data only assembly	23	——	18	169
CP034333.1 reference	39	18	——	
Hybrid assembly	184	169	185	——

Supplemental Table 5. Genome-wide pairwise comparison statistics from MuMmer4 for the MinION data only assemblies versus the hybrid and Illumina data-polished MinION assemblies for each species, related to STAR Methods.

		Genome-wide nucleotide-level pairwise identity where aligned	No. gSNPS	No. gIndels
B. malayi	**MinION data only assembly vs. hybrid**	99.57%	35,115	36,140
	MinION data only assembly vs. MinION assembly polished with Illumina data	99.89%	7,300	15,338
T. trichiura	**MinION data only assembly vs. hybrid**	99.22%	150,681	9,434
	MinION data only assembly vs. MinION assembly polished with Illumina data	99.86%	23,322	4,413
A. caninum	**MinION data only assembly vs. hybrid**	99.04%	299,510	49,610
	MinION data only assembly vs. MinION assembly polished with Illumina data	99.74%	147,516	45,902