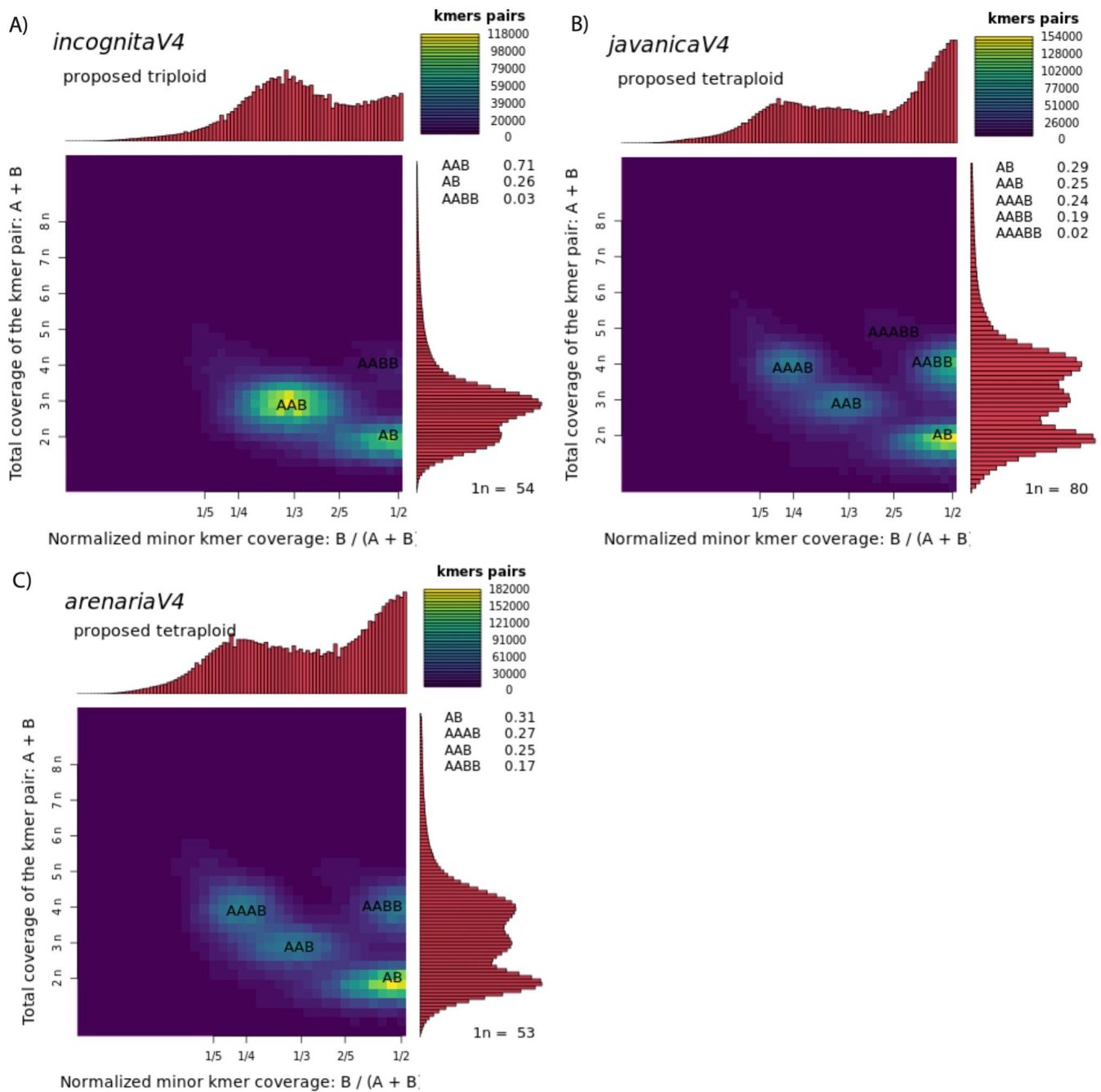


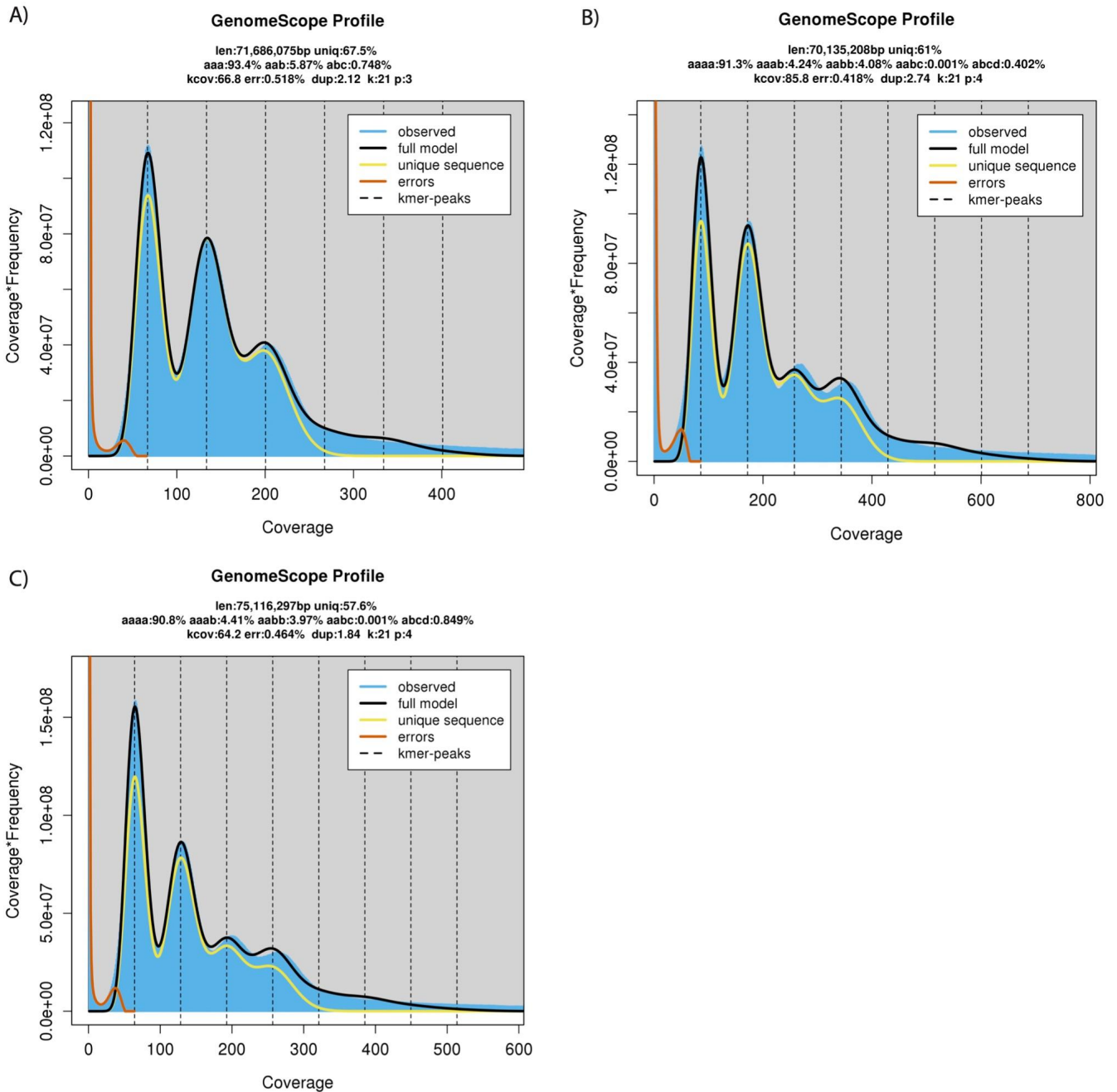
Supplementary information

Supplementary Figure. 1: k-mer estimation of genome ploidy



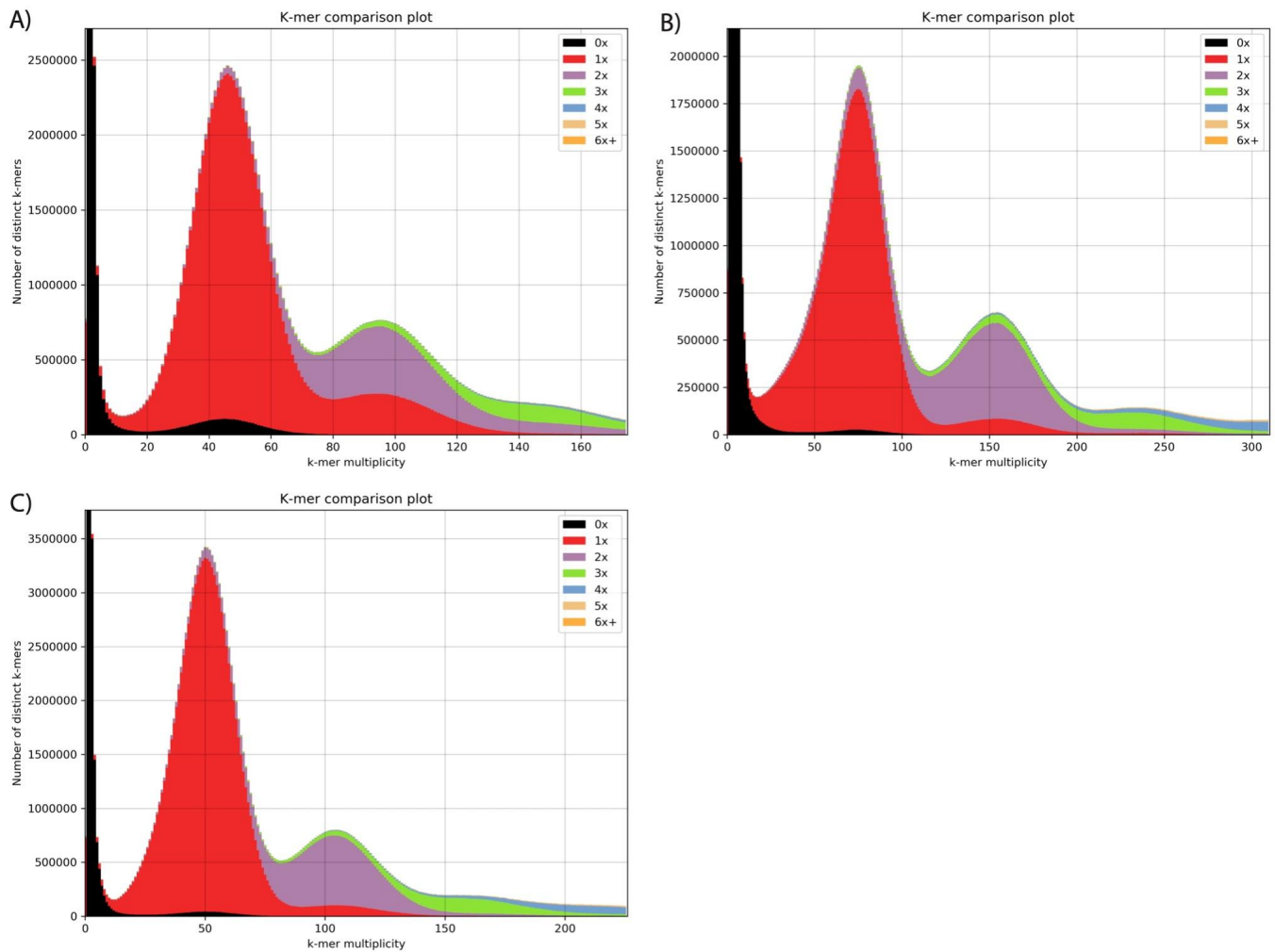
Smudgeplot¹ k-mer distribution analysis of the Illumina reads used for polishing the contigs predicted a triploid (3n 'AAB') genome structure for *M. incognita* (A) and tetraploid (4n 'AABB') genome structures for both *M. javanica* (B) and *M. arenaria* (C).

Supplementary Figure. 2: k-mer estimation of genome size and heterozygosity



Estimation of genome size and nucleotide divergence between homoeologous genome copies using GenomeScope2¹. **(A)** Setting the ploidy at 3n in *M. incognita* yielded an estimated haploid genome size of ca. 71.69Mb (and thus a total genome size of ca 215Mb), with an average nucleotide divergence between homoeologous genome copies of 6.6%. **(B)** Setting the ploidy at 4n in *M. javanica* yielded an estimated haploid genome size of ca. 70.14 Mb (and total genome size of 280.6 Mb) with an average nucleotide divergence between genome copies of 8.7%. **(C)** Setting the ploidy at 4n in *M. arenaria* yielded an estimated haploid genome size of ca. 75.12 Mb (total genome size of ca. 300.5 Mb) with an average nucleotide divergence between genome copies of 9.2%.

Supplementary Figure 3 k-mer estimation of genome assembly completeness

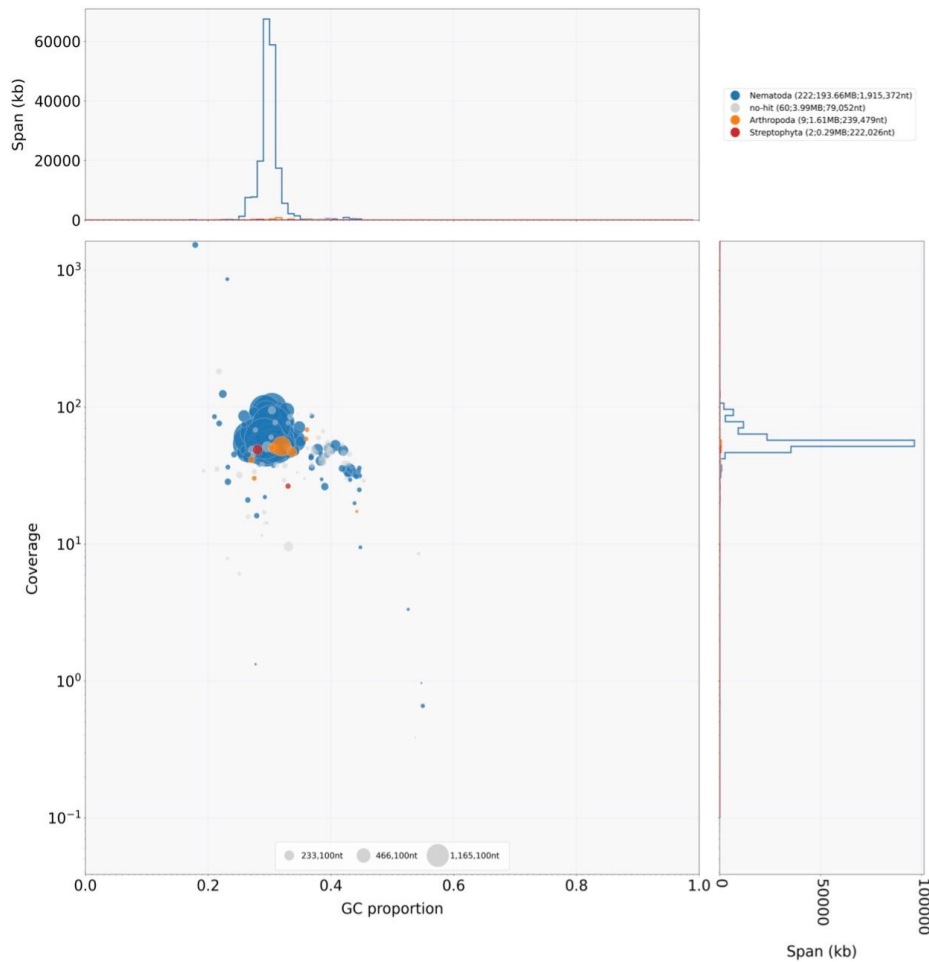


KAT² comparative analysis of the distribution of Illumina reads k-mers in the *M. incognita* (A), *M. javanica* (B), and *M. arenaria* (C) genome assemblies. Very few k-mers in the reads were not represented in the assembly (0X black curve), suggesting almost all the information present in the reads was retrieved in the assemblies. This observation and the congruence between genome assembly sizes and those estimated via flow cytometry³ suggest complete genomes with homoeologous genome copies having been mostly unzipped during assembly.

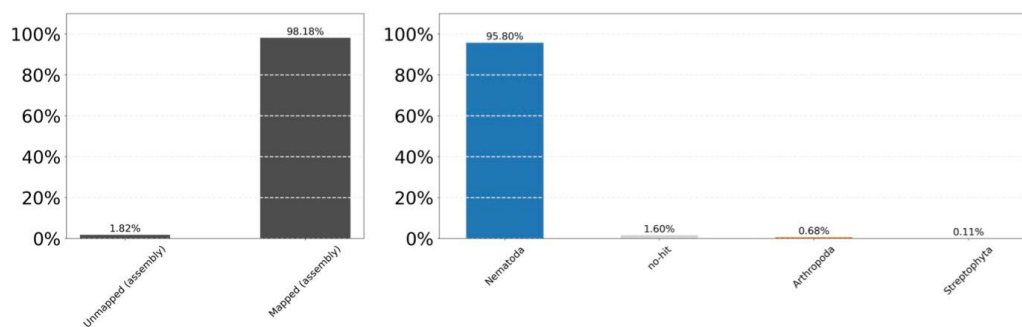
Supplementary Figure 4: genome contamination assessment with blobtools

1) *M. incognita*

A)



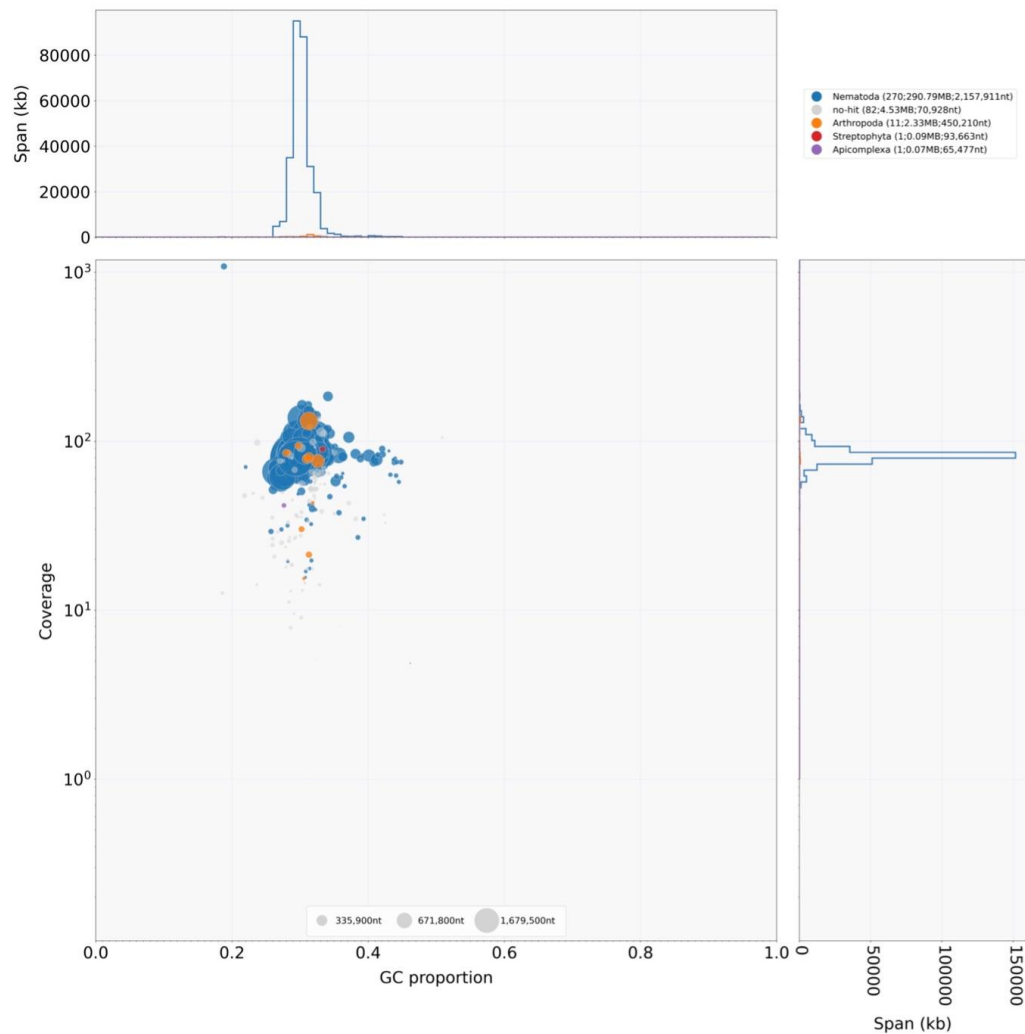
B)



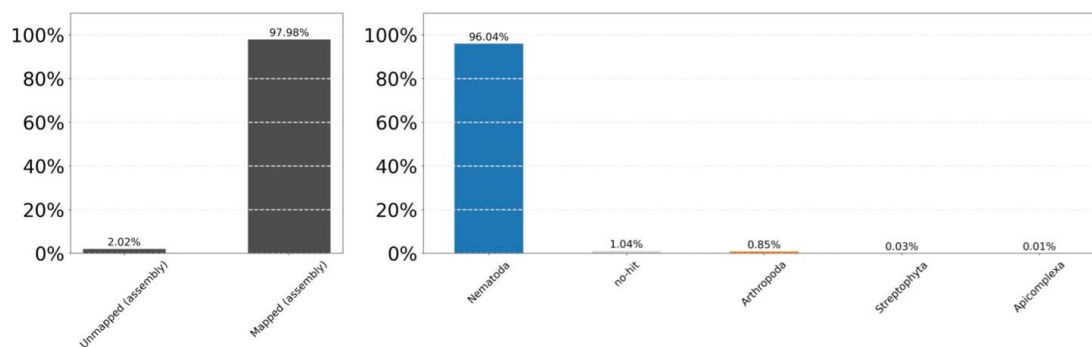
No evident trace of contamination could be identified. Two contigs with very high coverage corresponded to the mitochondrial genome of *M. incognita* and were removed from the final nuclear genome assembly.

2) *M. javanica*

A)



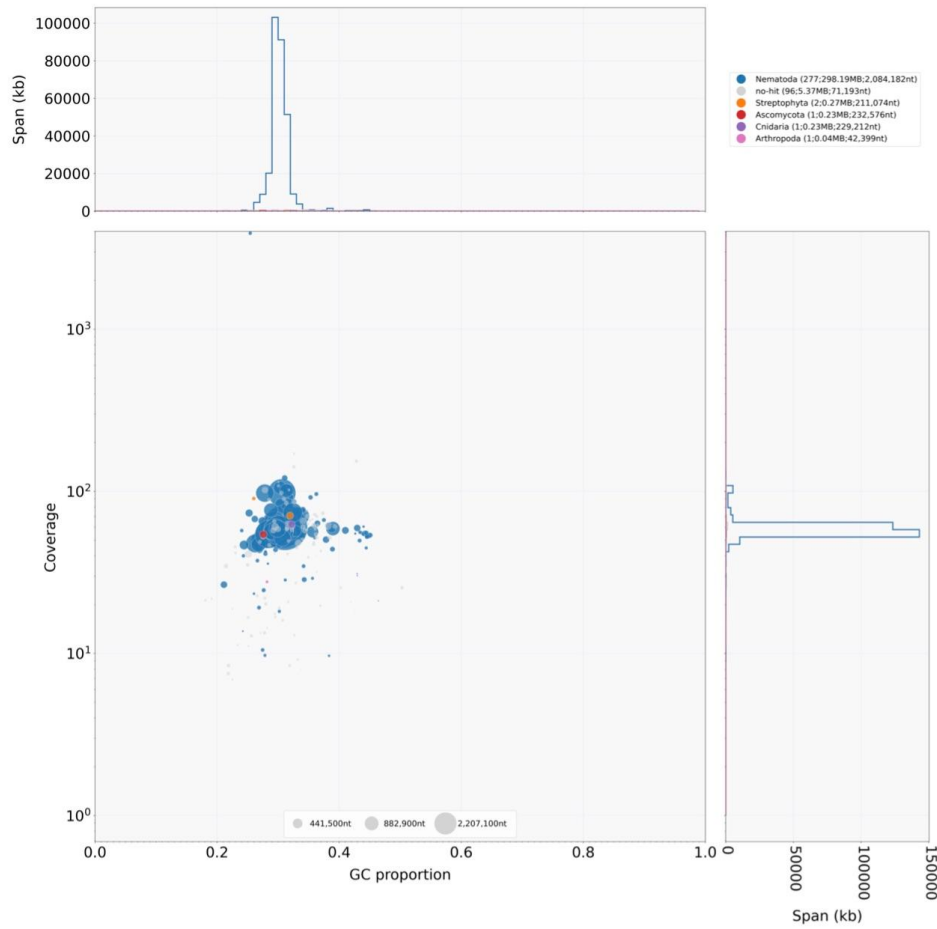
B)



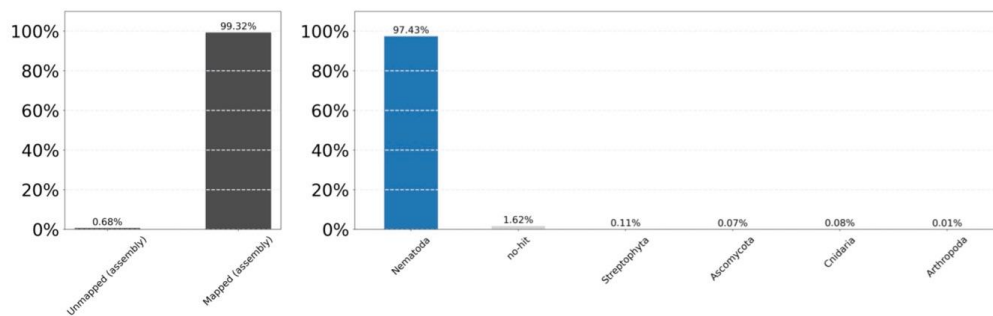
The majority of the contigs had approximately the same coverage, and one contig identified with a high coverage was annotated as the mitochondrial genome of *M. javanica* and removed from the assembly. A low percentage of the contigs was mapped as Arthropoda (0.85%), but this can be the result of a lack of information on the NCBI database.

3) *M. arenaria*

A)



B)



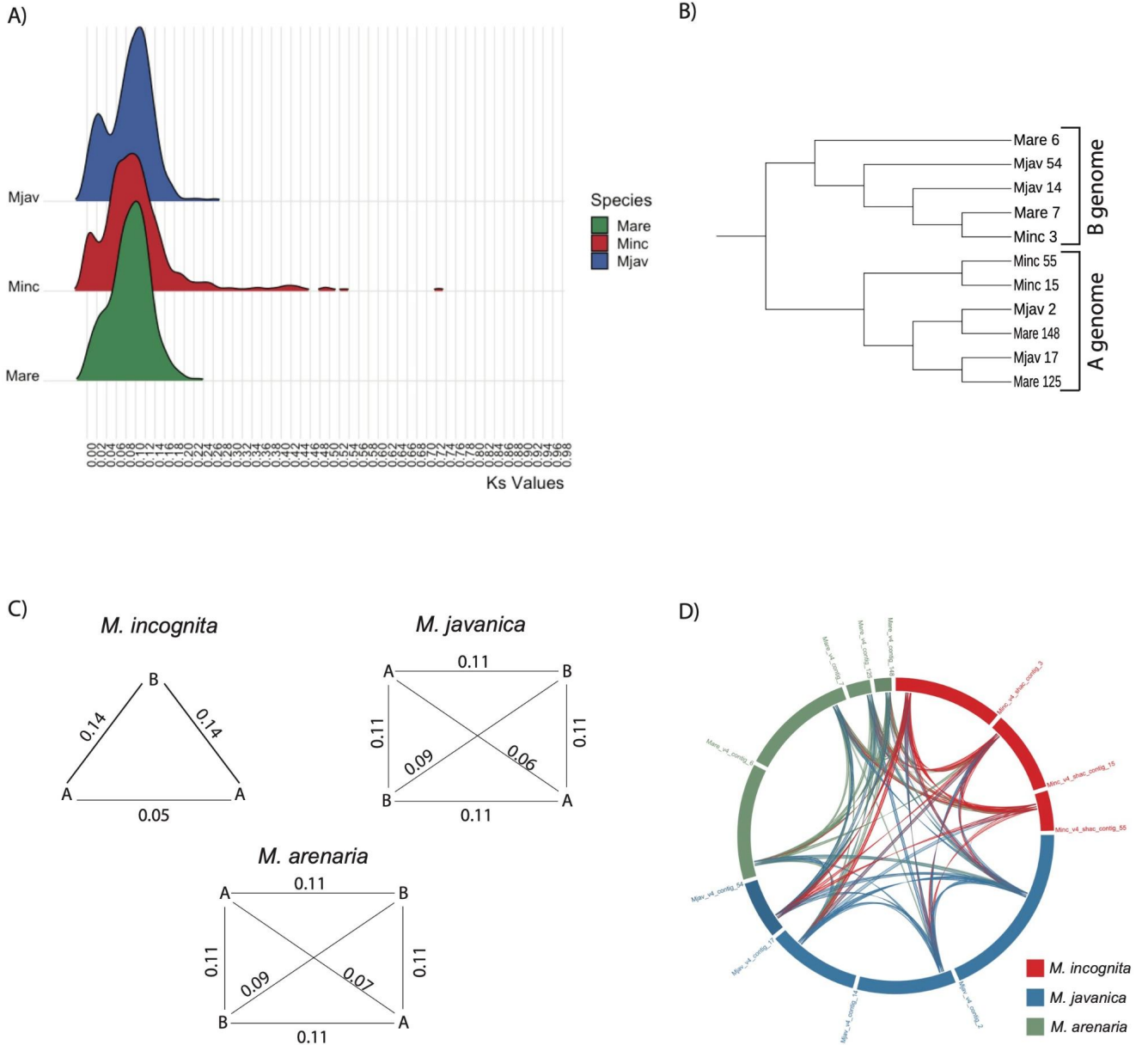
One contig was identified with a high coverage ($> 10^3$) compared to the others. This contig corresponded to the mitochondrial genome of *M. arenaria* and was eliminated from the final assembly. No contamination was observed in the final assembly.

Supplementary Table 1: Statistics of EuGene predictions

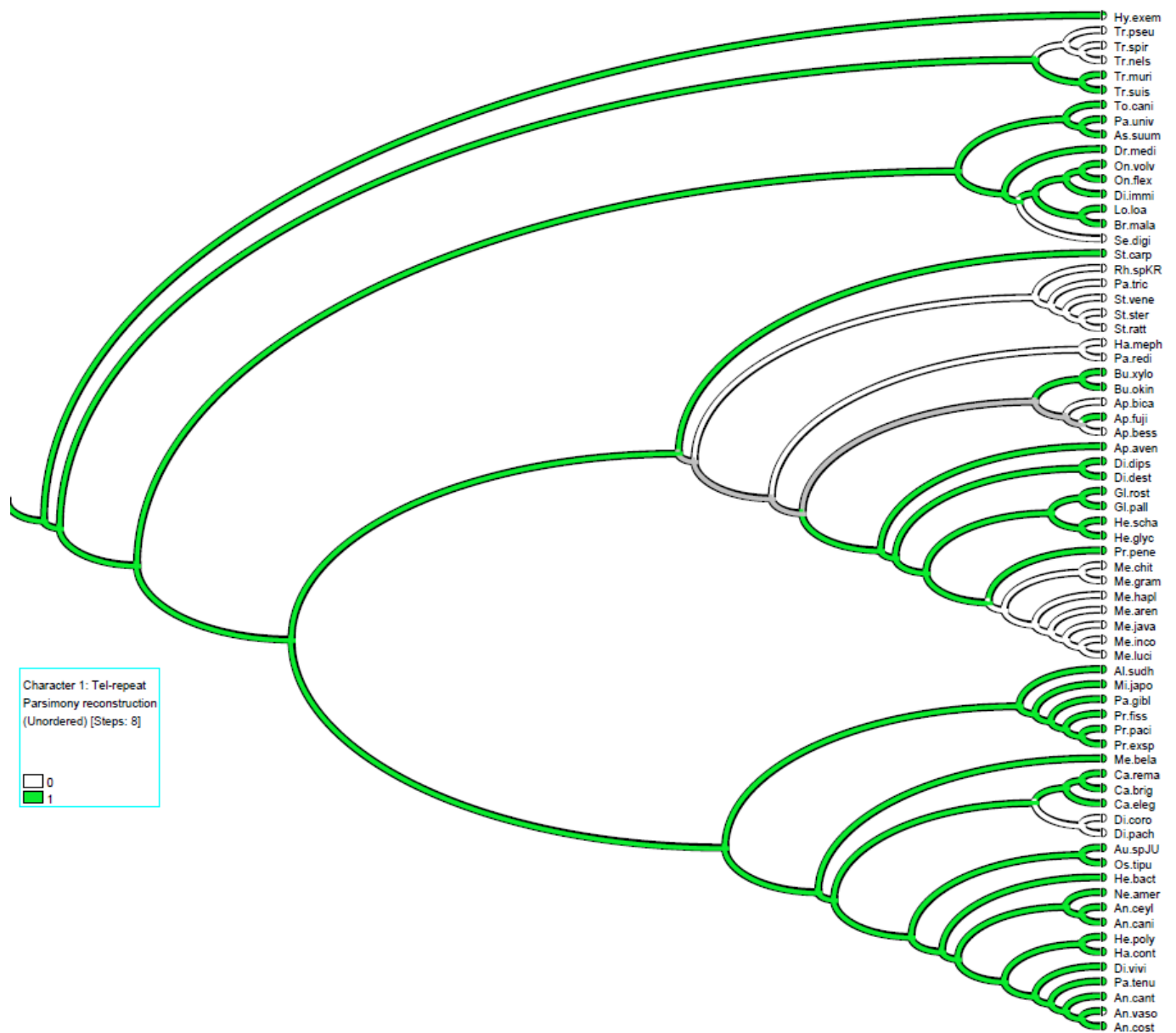
Feature / species	<i>Minc</i>	<i>Mjav</i>	<i>Mare</i>	<i>Mruci</i>
Predicted genes	54,816	64,689	63,434	52,935
Protein-coding	52,182	61,060	59,500	49,988
Avg. intergenic distance (coding genes) in bp	1,442.7	2,563.3	2,745.6	1,685.4
Avg. length protein-coding genes (bp)	2,254	2,102	2,122	2,333
Percent of the genome coding	23.6%	19.2%	18.5%	22.3%
GC% in coding regions	34.43%	34.63%	34.63%	34.56%
Percent of genes with introns	75%	82%	83%	79%
Avg. introns / gene	4.4	4.7	4.8	4.7
Avg. intron length	216.4	203.8	203.1	230.6
Percent of noncanonical (GC) splice donor	1.2%	1.0%	1.0%	1.3%
Non protein-coding genes	2,634	3,692	3,934	2,947
GC% of non-coding genes	38.6%	39.97%	40.9%	38.3%
Percent of genome masked by RED ¹	15.8%	23.8%	24.3%	16.4%

1: regions masked by EuGene⁴ using RED⁵ before gene prediction because they are repetitive with no evidence for transcription.

Supplementary Figure 5: Subgenome assignment based on gene collinearity and Ks.

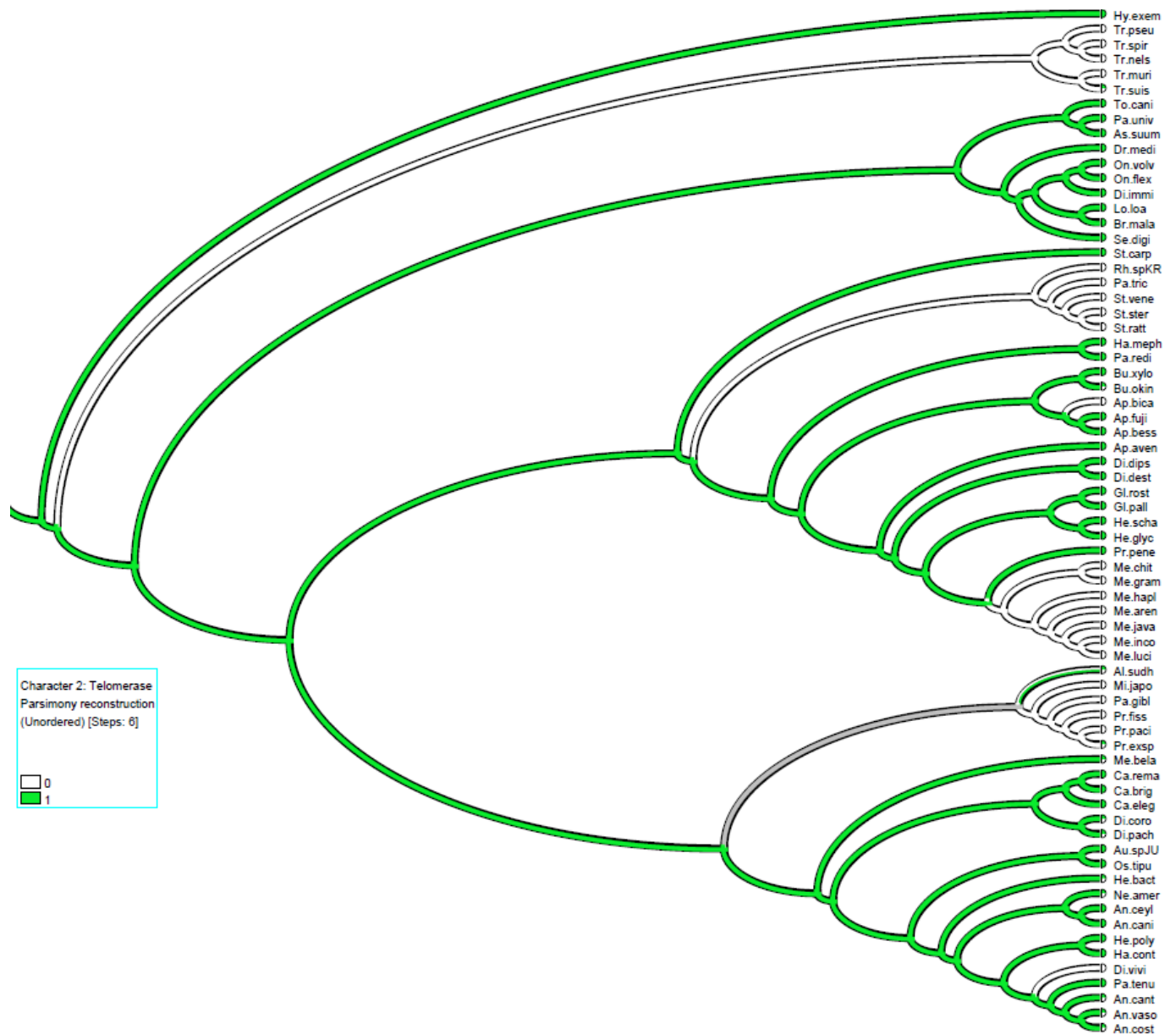


Supplementary Figure 6: telomeric repeat ancestral state reconstruction



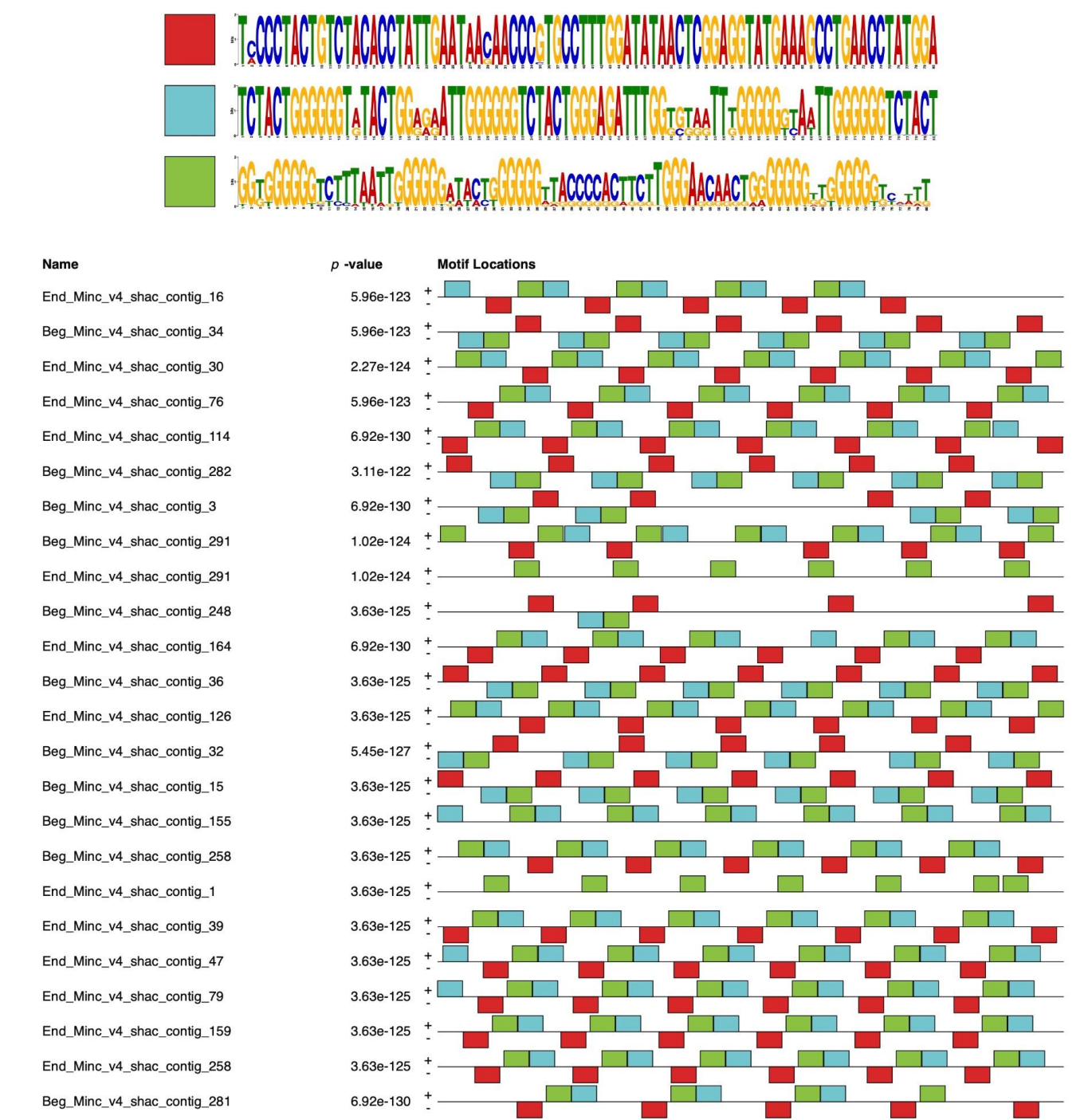
Based on presence / absence of the (TTAGGC) n telomeric repeat of *C. elegans* in nematode genomes, ancestral states were reconstructed using parsimony in Mesquite⁶. Green means (TTAGGC) n present and white absent, gray means presence and absence are equi-parsimonious. In the tardigrade *Hypsibius dujardini*, another simple repeat, (GATGGGTTTT) n , was described as a candidate telomeric repeat⁷.

Supplementary Figure 7: telomerase ancestral state reconstruction



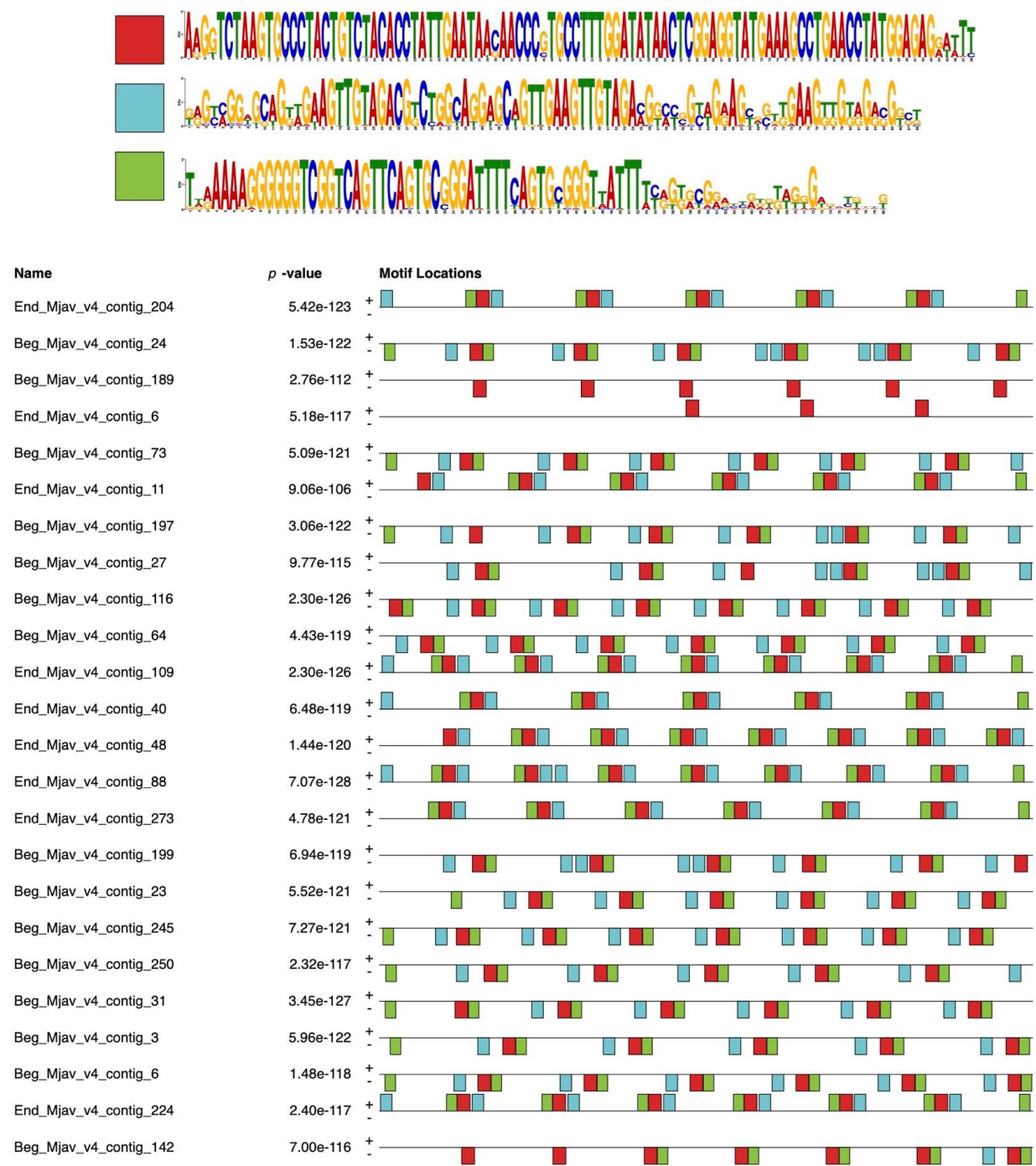
Based on presence / absence of the telomerase enzyme in nematode proteomes or telomerase gene in the genomes, ancestral states were reconstructed using parsimony in Mesquite⁶. Green means telomerase present and white absent.

Supplementary Figure 8: enriched motif at *M. incognita* contig ends



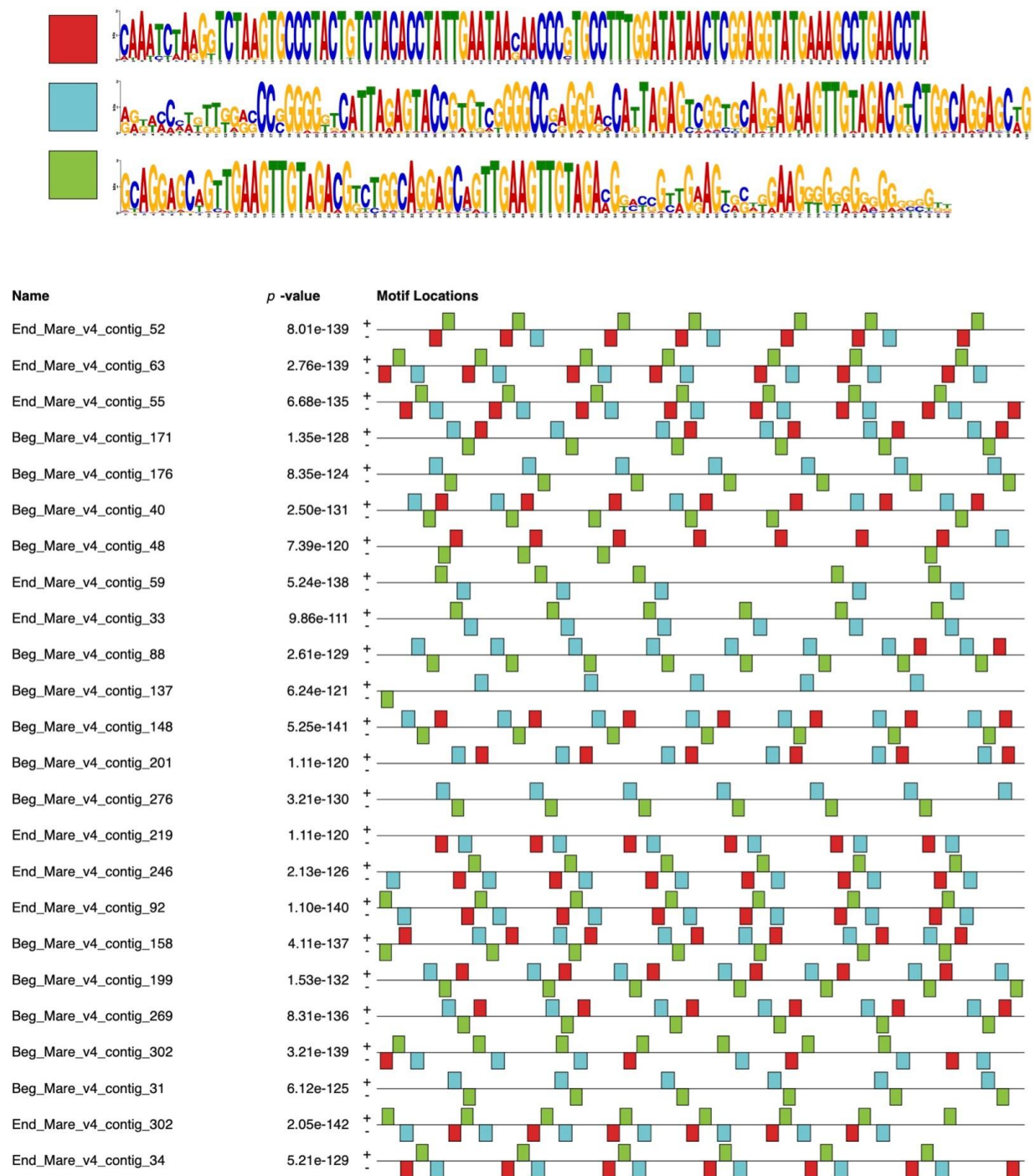
Enriched repeated motifs by increasing p-value and their distribution on contig extremities (first and last 2kb) in the *M. incognita* genome. Red: *Minc* motif-1 (MiM1), blue motif-2 (MiM2), and green: motif-3 (MiM3).

Supplementary Figure 9: enriched motif at *M. javanica* contig ends



Enriched repeated motifs by increasing p-value and their distribution on contig extremities (first and last 5kb) in the *M. javanica* genome. Red: *Mjav* motif-1 (MjM1), blue motif-2 (MjM2), and green: motif-3 (MjM3).

Supplementary Figure 10: enriched motif at *M. arenaria* contig ends



Enriched repeated motifs by increasing p-value and their distribution on contig extremities (first and last 5kb) in the *M. arenaria* genome. Red: *Mare* motif-1 (MaM1), blue motif-2 (MaM2), and green: motif-3 (MaM3).

Supplementary Figure 11: consensus of the enriched repeat at *M. incognita* contig extremities



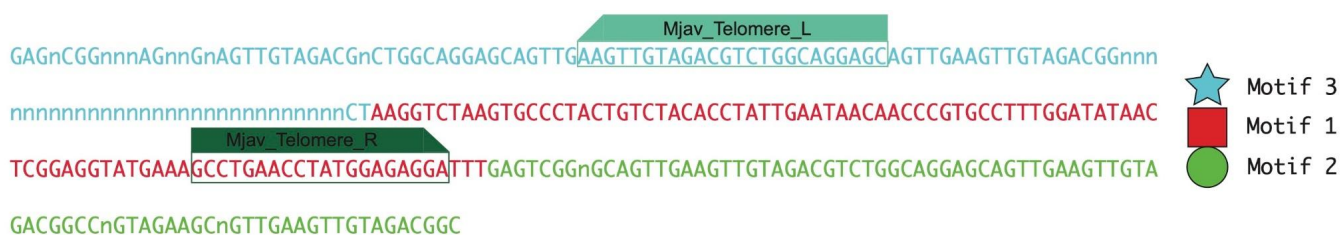
Positions of the primers used for PCR and FISH experiments are indicated as Minc_Telomere-R and Minc Telomere L

GGAGTGCATATTTTGTCTGGGGGTCTACTGGGGGTCTATTGGGGGTCTACTGGGGGTGTACTGGAGAATT
GGGGGTCTACTGGGAGATTTGGGGGTGGGGGTCTATTGGGGGTCTACTTGGGTGGGGGTCTTTAATTGG
GGGATACTGGGGGTACCCCACTTCTTGGGAACAACCTGGGGGGTTGGGGGTCTATTGGGGGTCTAGTTTC
CCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTCGGAGGTATGAAAGCCTGAAC
CTATGGACGAATTG

(in italics: positions of the primers)

Multiple sequence alignment of the repeated unit made of motif-2 (blue), motif-3 (green), and motif-1 (red) as defined in supplementary Figure 7 allowed deducing a ca. 250 - 300 bp consensus sequence.

Supplementary Figure 12: consensus of the enriched repeat at *M. javanica* contigs extremities

[illegible]

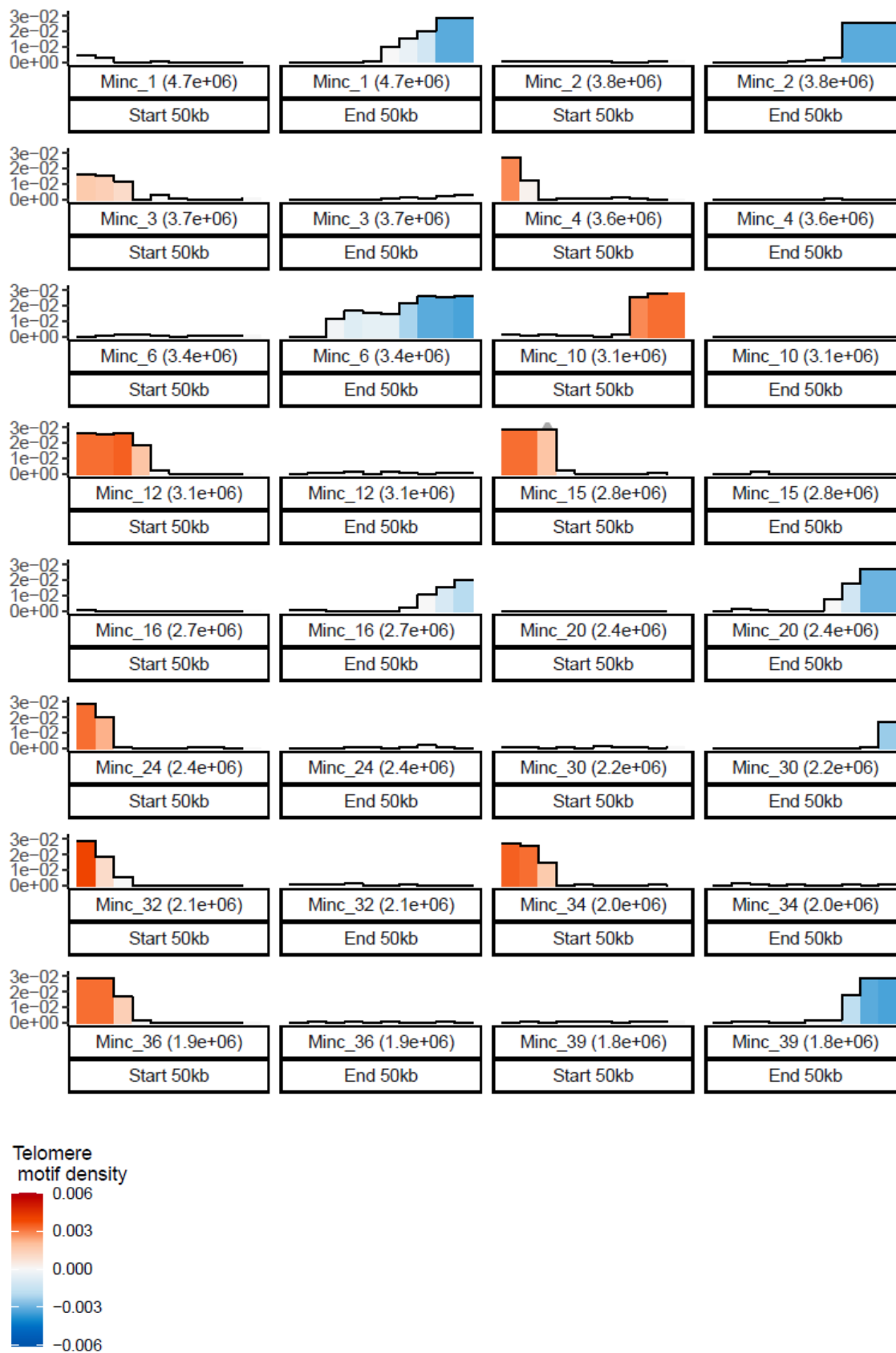
(in italics: positions of the primers)

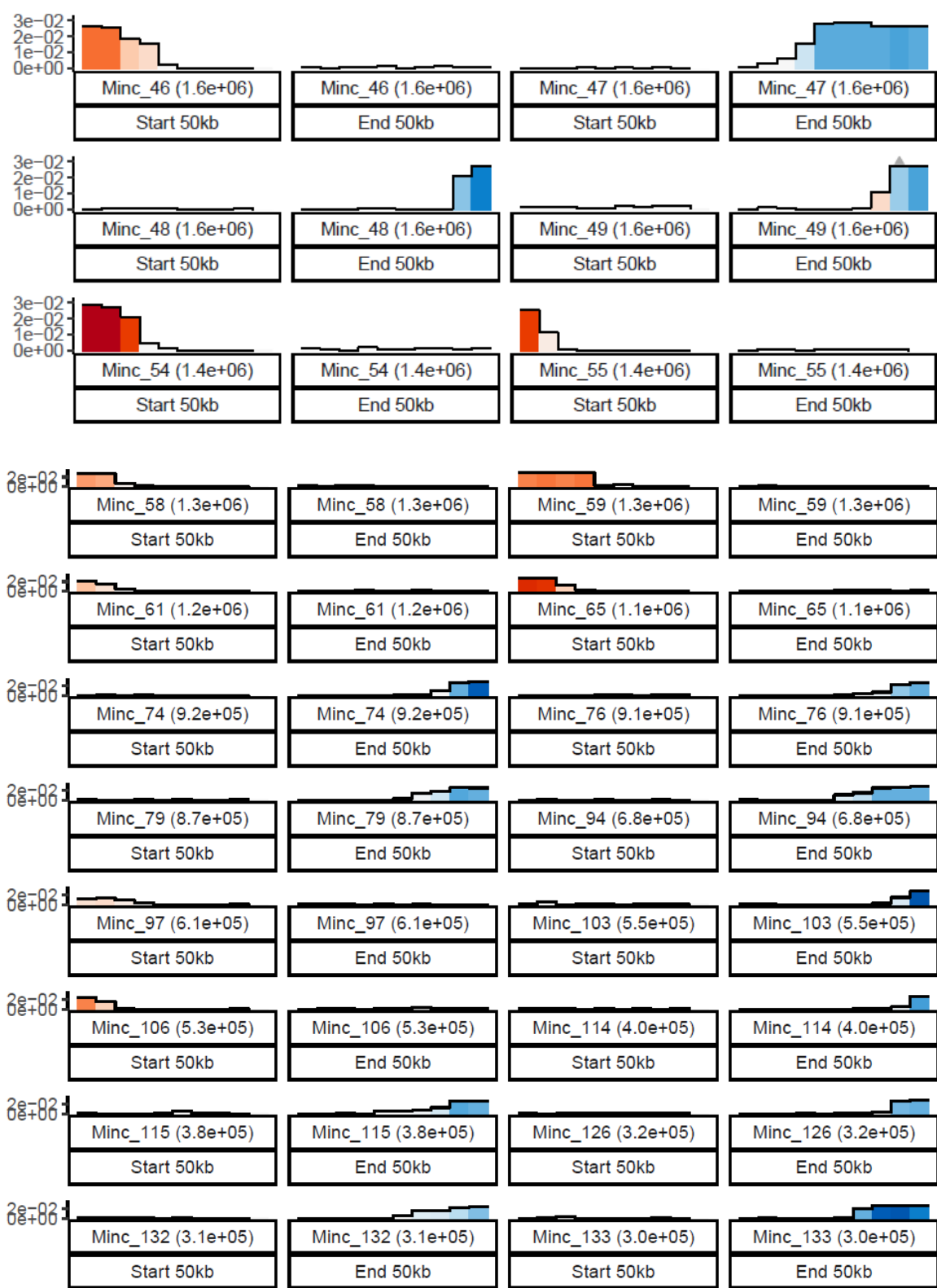
Supplementary Figure 13: consensus of the enriched repeat at *M. arenaria* contig extremities

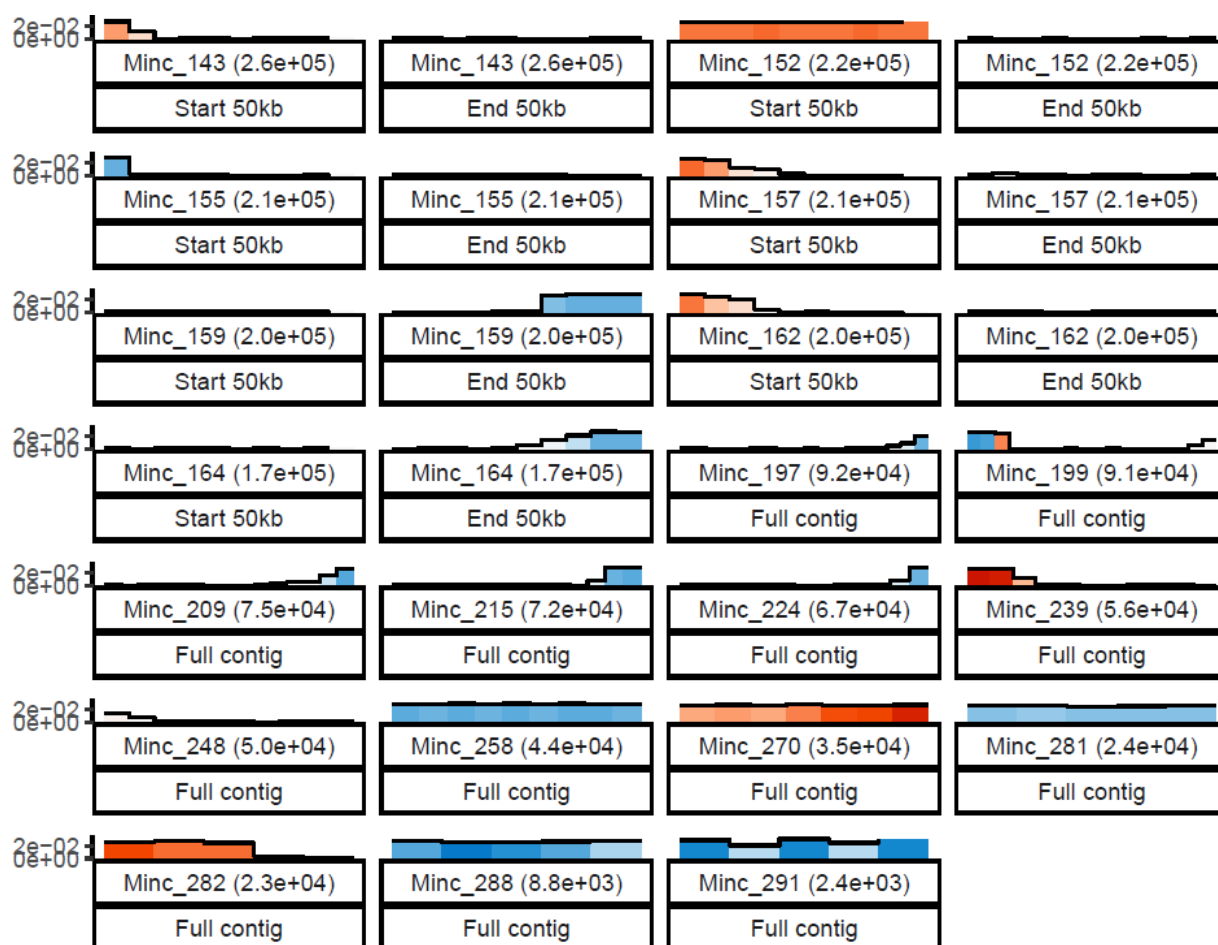


(in italics: positions of the primers)

Supplementary Figure 14: distribution of repeat patterns and G4-quadruplex on *M. incognita* contig 50kb extremities

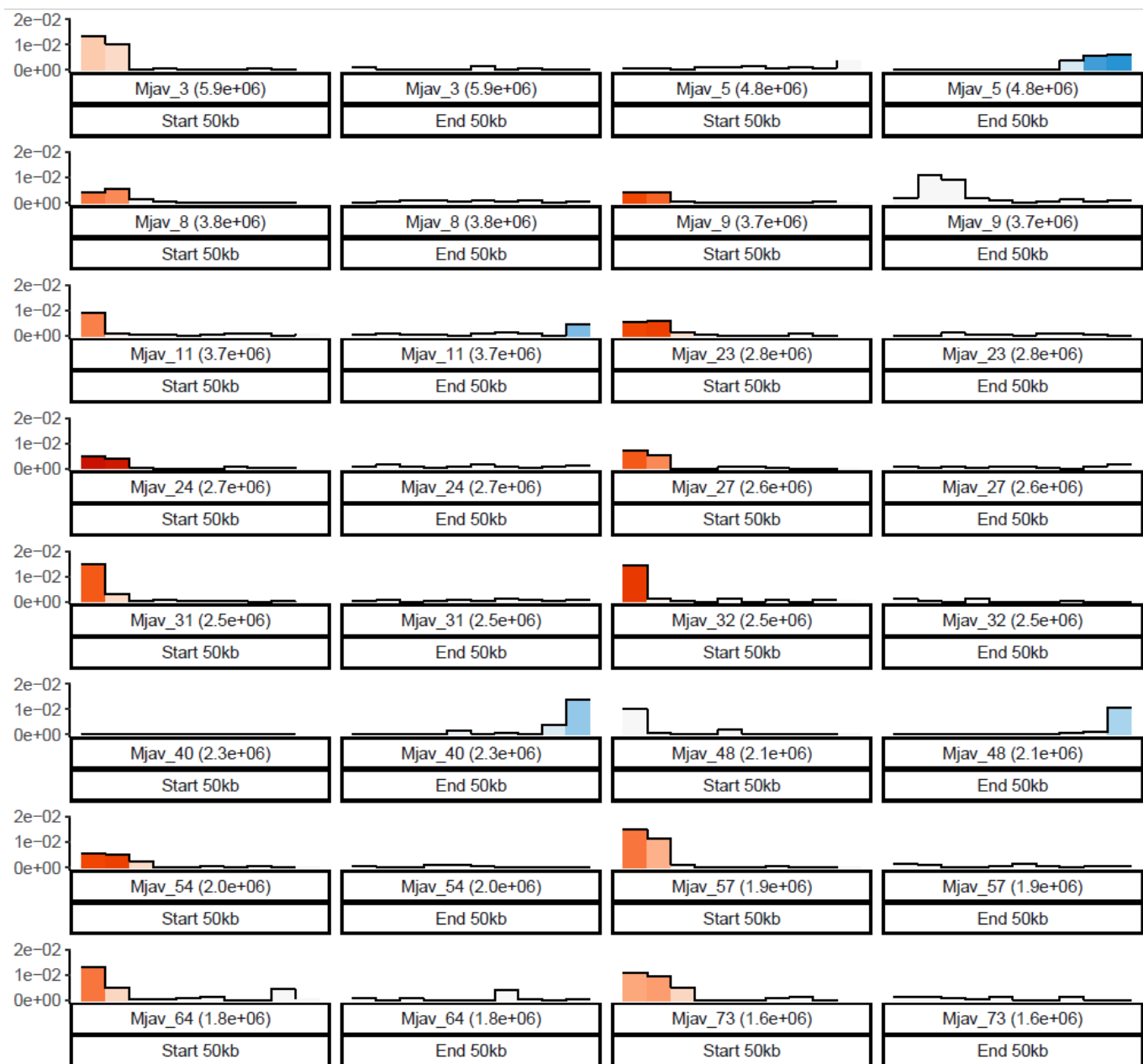


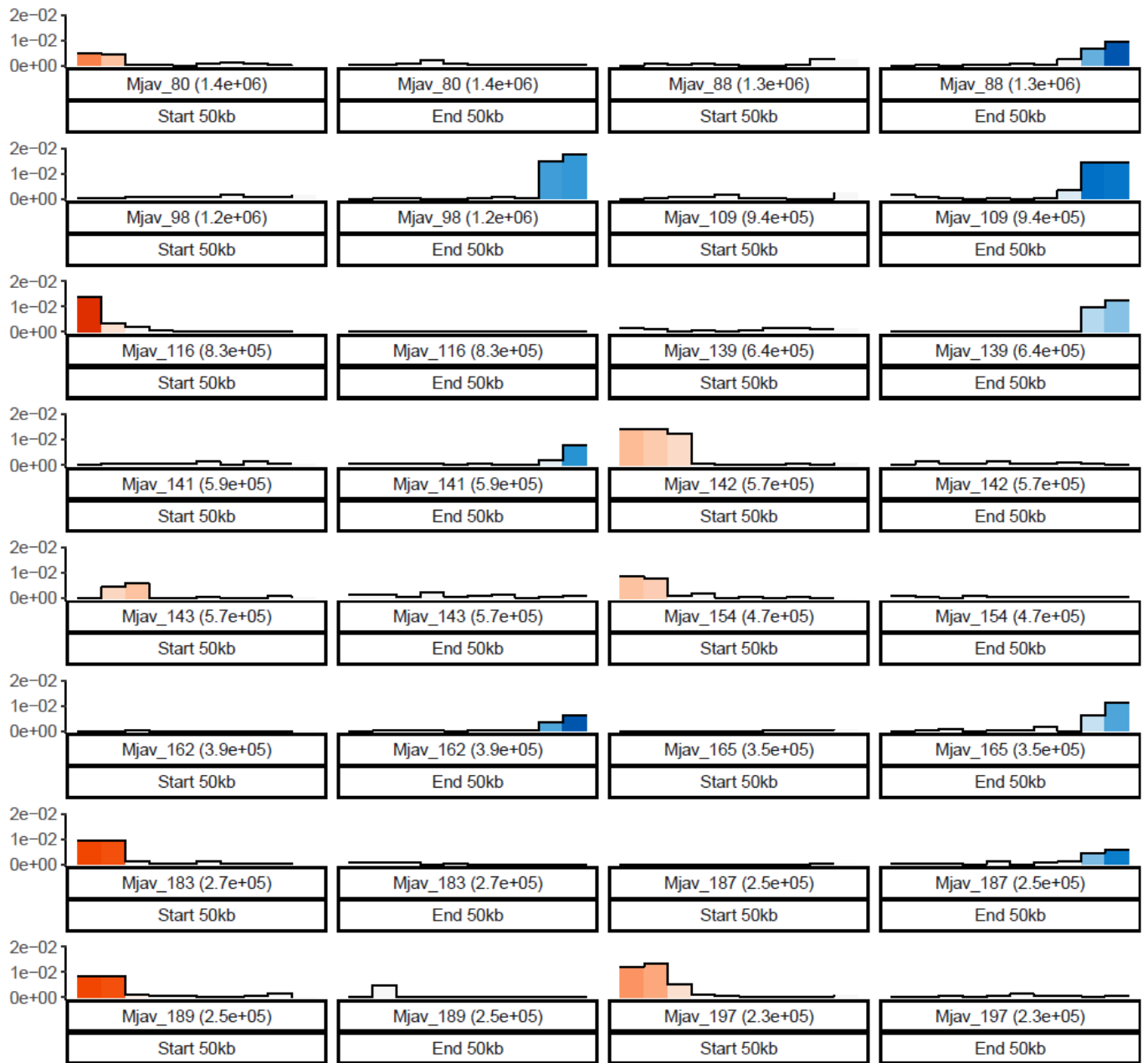




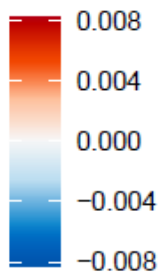
Density of repeat patterns and G4-quadruplex along *M. incognita* contig extremities. The density of *Minc* repeat units is represented by a color gradient with positive values (red) indicating a density in the sense strand and negative values (blue) indicating a density on the reverse complement strand. Gray triangles above bars indicate regions of the contigs where less than 100% of the repeat patterns are on the same strand. The heights of bars in the histogram represent the density of G4-quadruplex forming regions. The first and last 50kb of each contig containing at least 3 repeats are represented and the values are per 5kb windows on the genome. As a complement to Figure 3, a supplementary figure online represents the distribution of *Minc* repeats and G4s on the whole contig length in 100kb windows (<https://doi.org/10.57745/1WDPE4>). Source data to produce this figure is provided as a Source Data file.

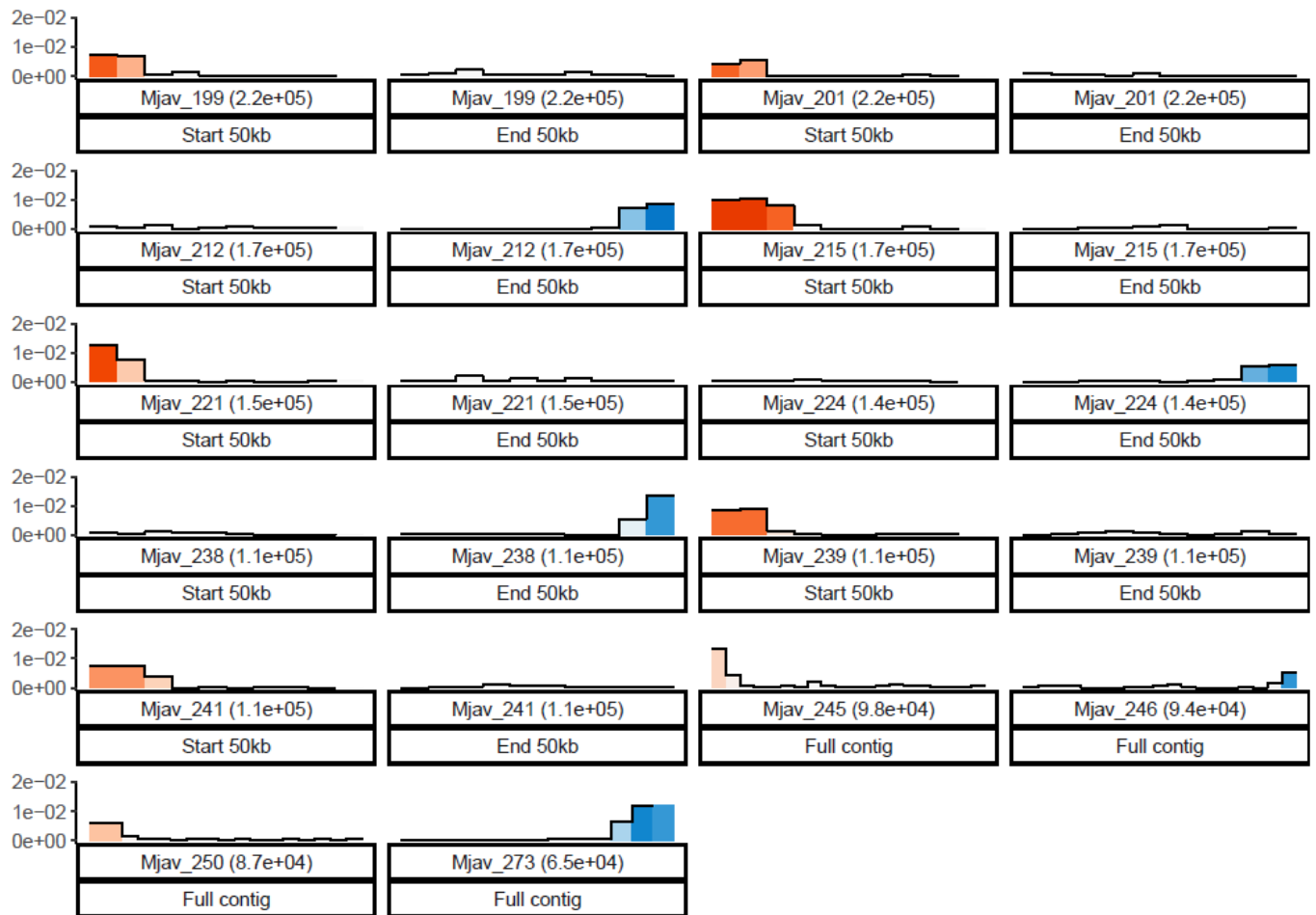
Supplementary Figure 15: Distribution of repeat patterns and G4-quadruplex on *M. javanica* 50kb contig extremities





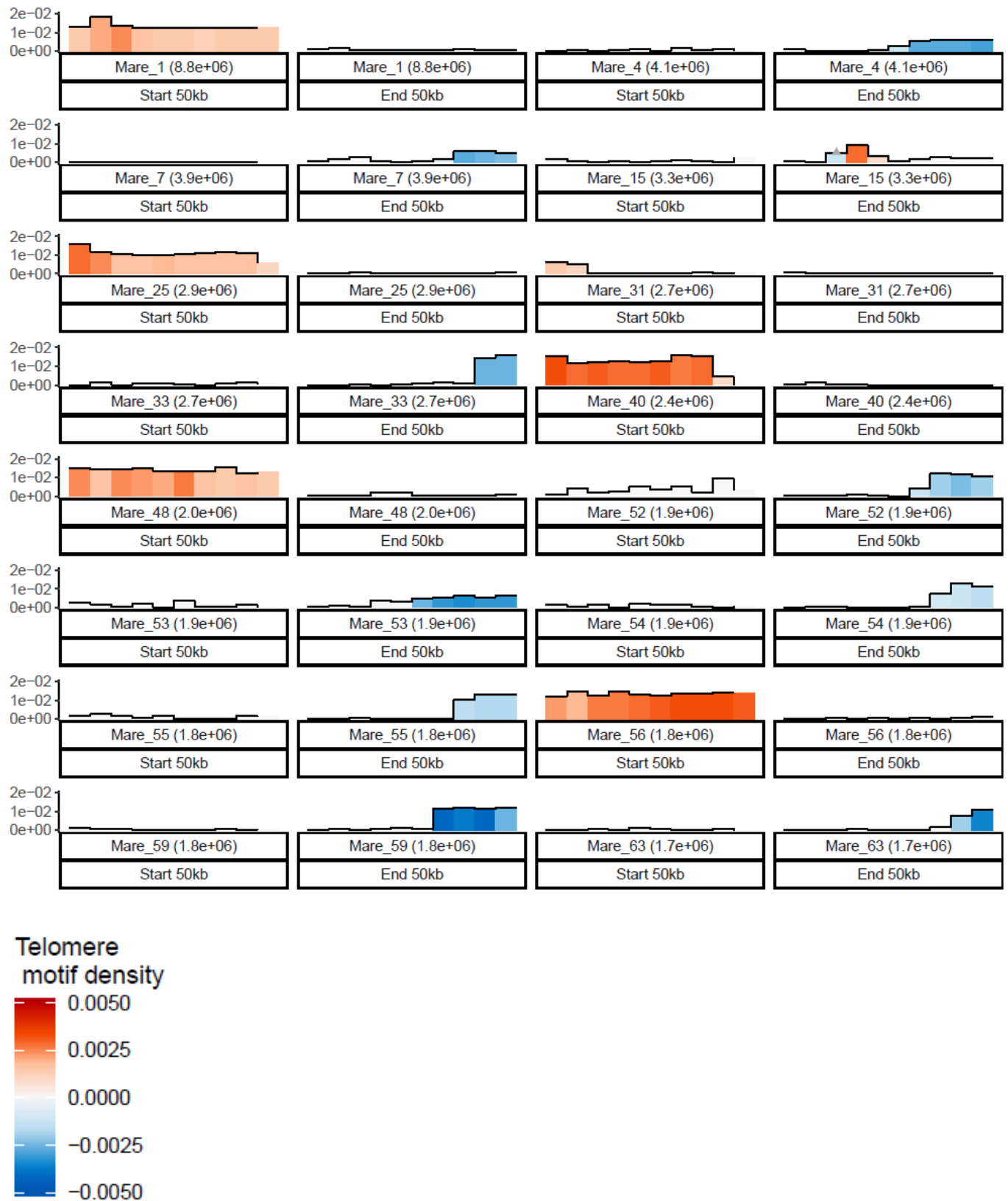
Telomere motif density

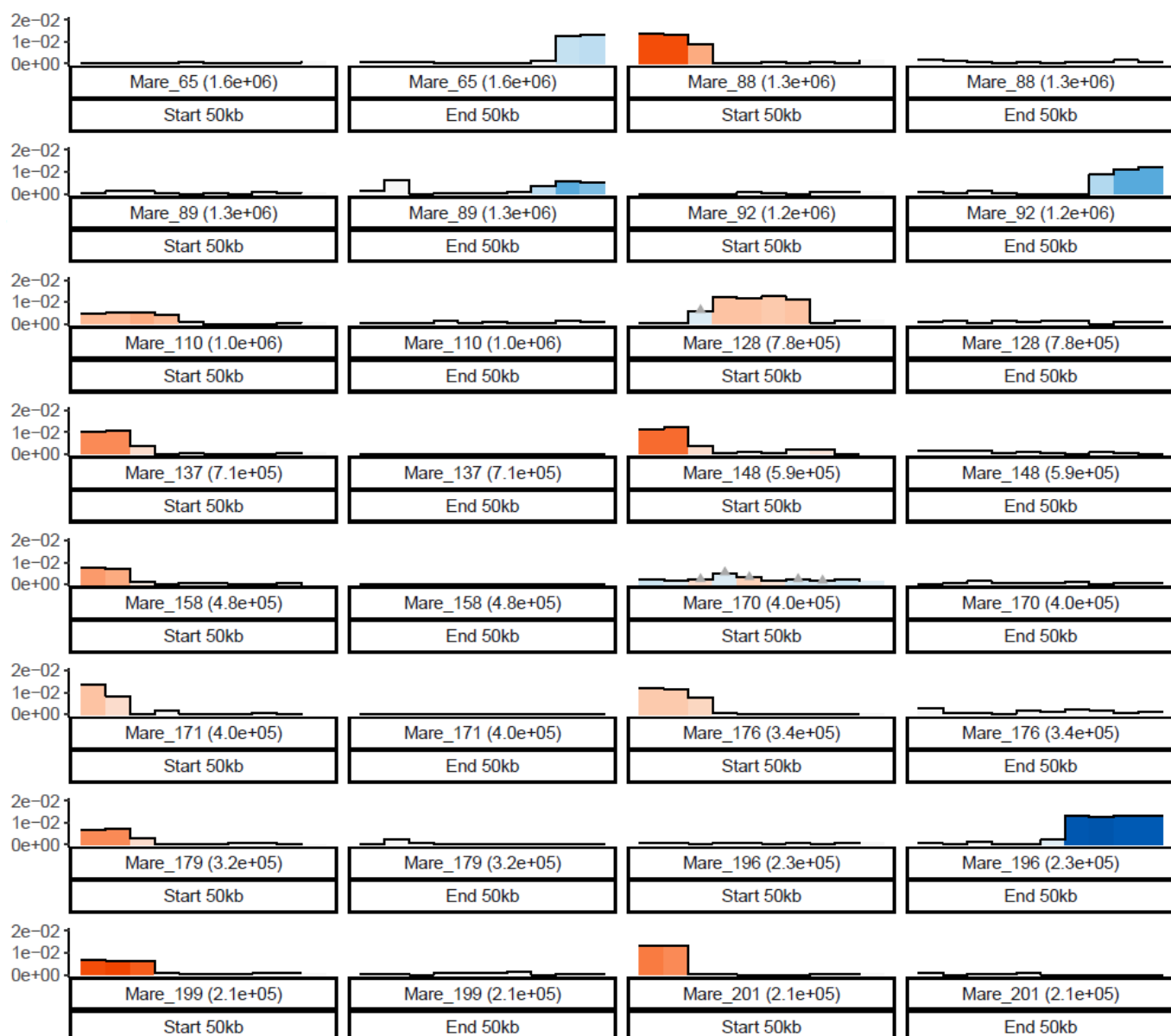


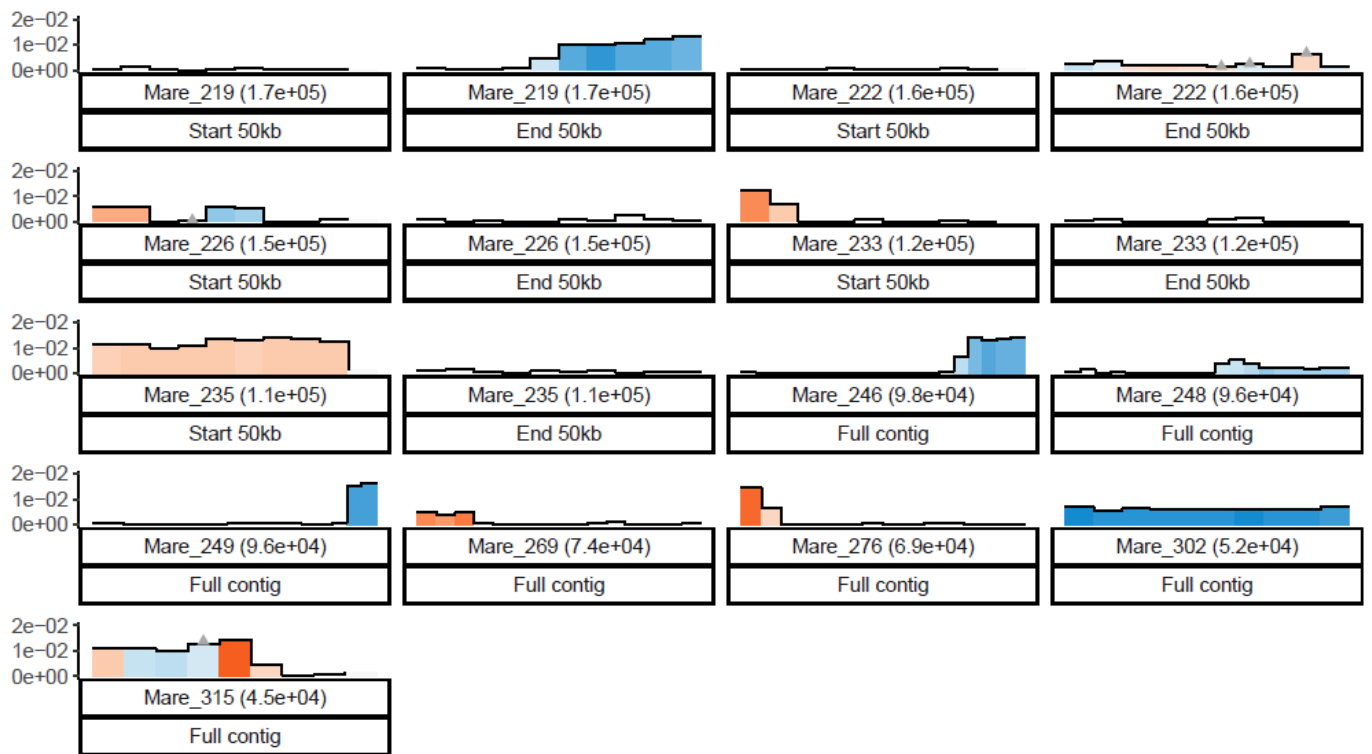


B) Density of repeat patterns and G4-quadruplex along *M. javanica* contig extremities. The density of *Mjav* repeat units is represented by a color gradient with positive values (red) indicating a density in the sense strand and negative values (blue) indicating a density on the reverse complement strand. Gray triangles above bars indicate regions of the contigs where less than 100% of the repeat patterns are on the same strand. The heights of bars in the histogram represent the density of G4-quadruplex forming regions. The first and last 50kb of each contig containing at least 3 repeats are represented and the values are per 5kb windows on the genome, contigs that do not contain repeats in their first or last 50kb are ignored. The distribution of *Mjav* repeats and G4s on the whole contig length in 100kb windows is available online at (<https://doi.org/10.57745/EIKAQ4>). Source data to produce this figure is provided as a Source Data file.

Supplementary Figure 16: Distribution of repeat patterns and G4-quadruplex on *M. arenaria* 50kb contig extremities







B) Density of repeat patterns and G4-quadruplex along *M. arenaria* contig extremities. The density of *Mare* repeat units is represented by a color gradient with positive values (red) indicating a density in the sense strand and negative values (blue) indicating a density on the reverse complement strand. Gray triangles above bars indicate regions of the contigs where less than 100% of the repeat patterns are on the same strand. The heights of bars in the histogram represent the density of G4-quadruplex forming regions. The first and last 50kb of each contig containing at least 3 repeats are represented and the values are per 5 kb windows on the genome, contigs that do not contain repeats in their first or last 50 kb are ignored. The distribution of *Mare* repeats and G4s on the whole contig length in 100 kb windows is available online at (<https://doi.org/10.57745/CY06YE>). Source data to produce this figure is provided as a Source Data file.

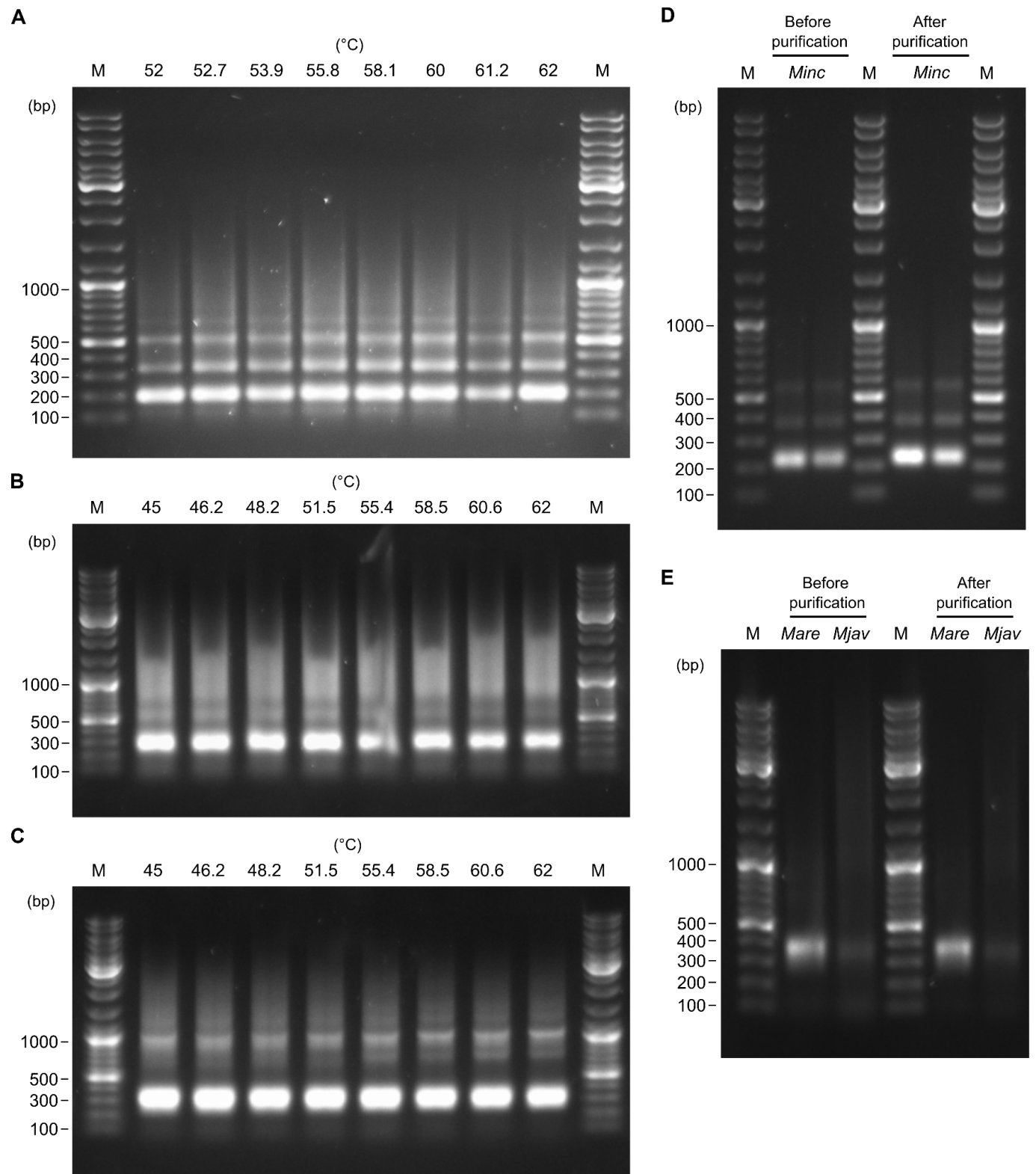
Supplementary Table 2: annotation of (retro)transposons and other repeats

The annotation of transposable elements and other repetitive regions on the genomes of *M. incognita*, *M. arenaria*, and *M. javanica* was performed using EDTA ⁸ version 2.1.0.

		<i>M. incognita</i>		<i>M. arenaria</i>		<i>M. javanica</i>	
Class	Type	Count	bpMasked	Count	bpMasked	Count	bpMasked
Retrotransposons							
LTR	Copia	25	8,783	444	185,968	290	134,146
	Gypsy	10216	5,548,981	12585	12,152,090	9704	8,940,100
	Unknown	11048	6,556,744	25891	11,396,311	11646	5,609,431
nonLTR	LINE_element	370	373,486	503	793,499	528	703,477
DNA Transposons							
TIR	CACTA	4296	896,156	5464	1,186,379	4370	1,113,443
	Mutator	30418	7,227,848	48371	12,631,831	54636	17,274,061
	PIF_Harbinger	2231	522,128	2513	696,,702	2096	531,815
	Tc1_Mariner	476	178,149	2128	522,,780	1963	524,585
	hAT	55062	11,913,136	94287	22,654,778	85261	19,036,926
	polinton	2561	2,988,992	1755	1,709,300	4784	5,536,271
nonTIR	helitron	51164	2,635,069	4369	4,726,375	4077	5,409,339
Unclassified / unknown repeats							
repeat region	repeat	27,789	8,705,800	45567	19,646,238	55511	19,620,486

Supplementary Figure 17. Gel electrophoresis of amplified telomere sequence.

Gradient PCR using primer pair and temperature range as listed in the Methods section in *M. incognita* (A), *M. javanica* (B), and *M. arenaria* (C). Labeled probe for telomere sequence using biotin-dUTP nucleotide in PCR reaction for *M. incognita* (D), and *M. javanica* as well as *M. arenaria* (E).



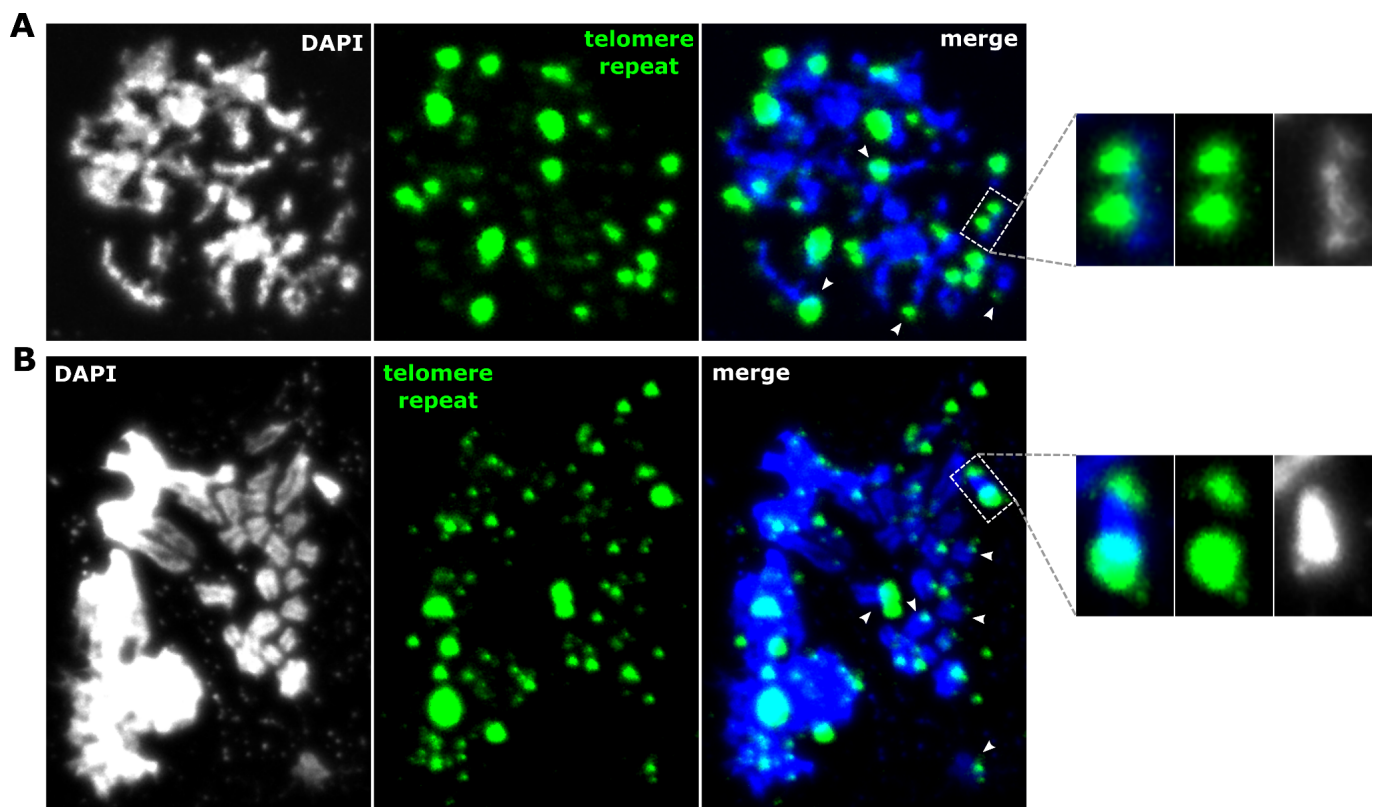
Supplementary Figure 18. FISH validation of telomeric localization of the repeat on *Minc* chromosomes

A high-resolution multi-panel image confirming specific FISH signal at telomeric position on *Minc* chromosomes as well as chromosomes and repeat arrays count in nuclei on several additional phases of the cell cycle <https://doi.org/10.57745/5SXZQH>

Supplementary Figure 19. Animated GIFs of FISH signal confirming specific localization at *Minc* chromosome extremities

Animated GIFs are available online at Recherche Data Gouv: <https://doi.org/10.57745/OWZLRN> with A) a 3D-reconstructed animation of *Minc* chromosomes with FISH signal and B) an animated GIF with an alternative display of the chromosome staining channel and fluorescence channel.

Supplementary Figure 20. Telomere localization in *M. javanica* (A) and *M. arenaria* (B).



Chromosomes with visible telomere positions at their end are marked with white arrows. Enlarged on the right are chromosomes that show telomeric localization at both ends.

Supplementary Table 3: Raw Illumina data used for genome polishing and k-mer analyses

Species	Raw sequencing library	Reads	Length	Accession numbers
<i>M. incognita</i>	BZU_AAOSDE_1_1_H732LBCX2.12BA290_clean.fastq	19,805,212	250	ERX10638998
<i>M. incognita</i>	BZU_AAOSDE_1_2_H732LBCX2.12BA290_clean.fastq	19,805,212	250	ERX10638998
<i>M. incognita</i>	BZU_AAOSDE_2_1_H732LBCX2.12BA290_clean.fastq	20,090,514	250	ERX10638999
<i>M. incognita</i>	BZU_AAOSDE_2_2_H732LBCX2.12BA290_clean.fastq	20,090,514	250	ERX10638999
<i>M. javanica</i>	BZU_ACOSDE_2_1_H52YLBCX2.12BA292_clean.fastq	64,456,448	250	ERX10639000
<i>M. javanica</i>	BZU_ACOSDE_2_2_H52YLBCX2.12BA292_clean.fastq	64,456,448	250	ERX10639000
<i>M. arenearia</i>	BZU_ABOSDE_1_1_H732LBCX2.12BA291_clean.fastq	27,049,894	250	ERX10639001
<i>M. arenearia</i>	BZU_ABOSDE_1_2_H732LBCX2.12BA291_clean.fastq	27,049,894	250	ERX10639001
<i>M. arenearia</i>	BZU_ABOSDE_2_1_H732LBCX2.12BA291_clean.fastq	27,421,721	250	ERX10639002
<i>M. arenearia</i>	BZU_ABOSDE_2_2_H732LBCX2.12BA291_clean.fastq	27,421,721	250	ERX10639002

Supplementary Table 4: Primer sequences used in telomere FISH experiments

	Left primer	Right primer	Annealing temperature (°C)
<i>M. incognita</i>	TCGTCCATAGGTTCAAGGCTT	GGGTCTACTTGGGTGGGG	62
<i>M. javanica</i>	AAGTTGTAGACGTCTGGCAGG	TCCTCTCCATAGGTTCAAGC	51.5
<i>M. arenaria</i>	GGGGTCATTAGAGTACCGTGTC	TTCAGGCTTTCATACCTCCGAG	55.4

References

1. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1–10 (2020).
2. Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J. & Clavijo, B. J. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics* **33**, 574–576 (2017).
3. Blanc-Mathieu, R. *et al.* Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics* **13**, e1006777 (2017).
4. Sallet, E., Gouzy, J. & Schiex, T. EuGene: An Automated Integrative Gene Finder for Eukaryotes and Prokaryotes. *Methods Mol. Biol.* **1962**, 97–120 (2019).
5. Girgis, H. Z. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics* **16**, 227 (2015).
6. Maddison, W. P. & Maddison, D. R. Mesquite: A modular system for evolutionary analysis. Version 3.01. (2014).
7. Yoshida, Y. *et al.* Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLOS Biology* **15**, e2002266 (2017).
8. Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).