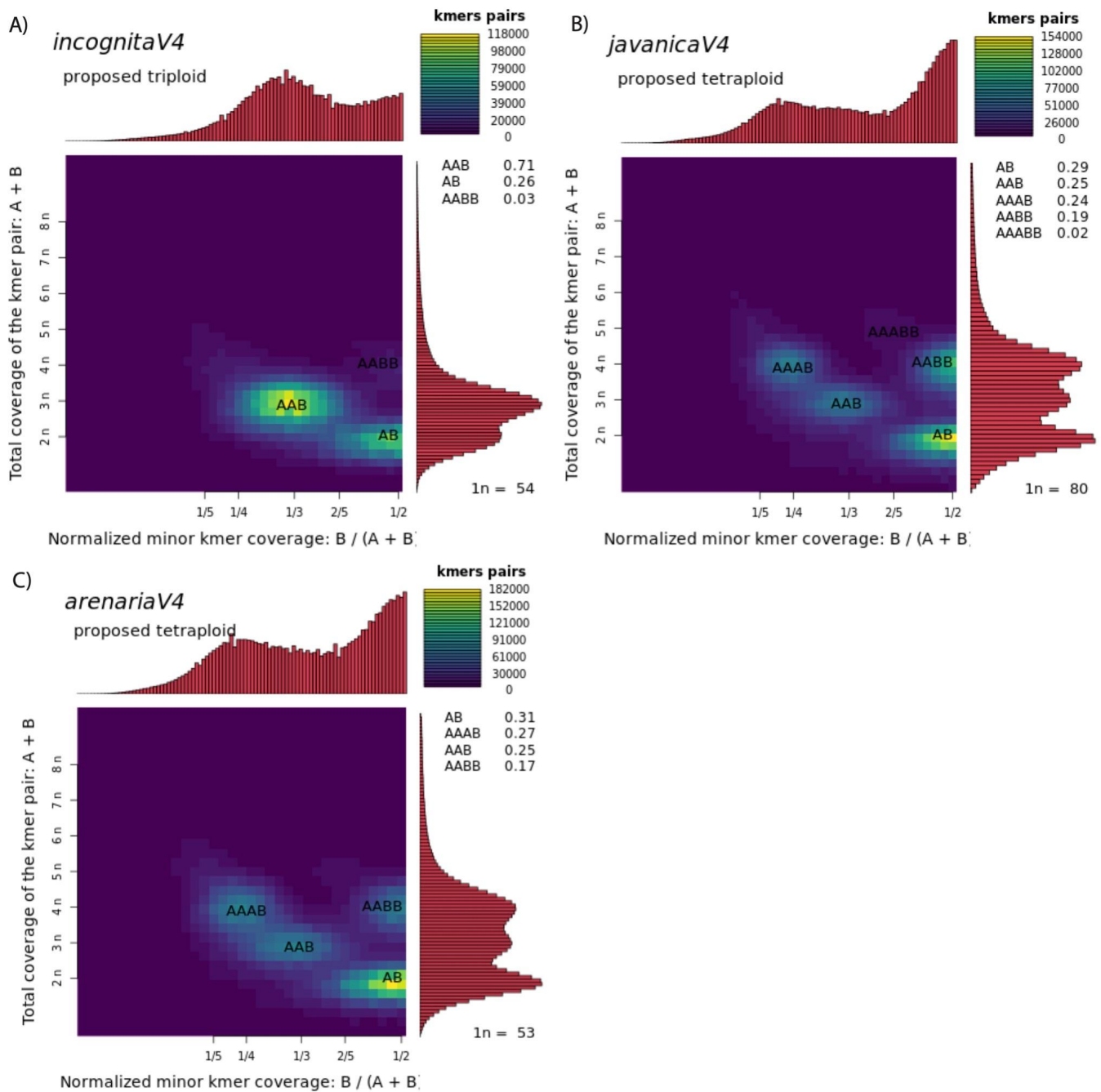


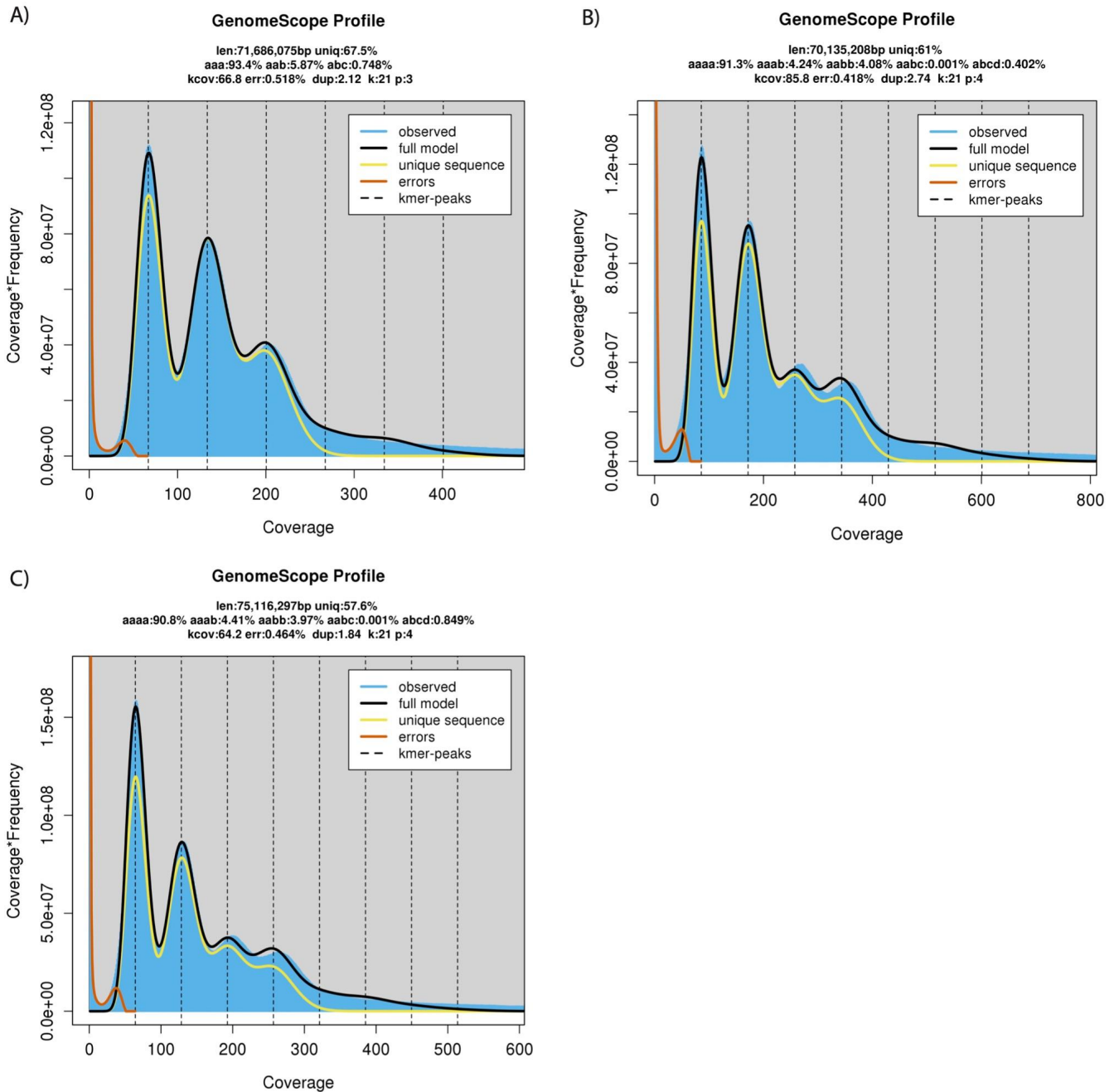
## Supplementary information

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



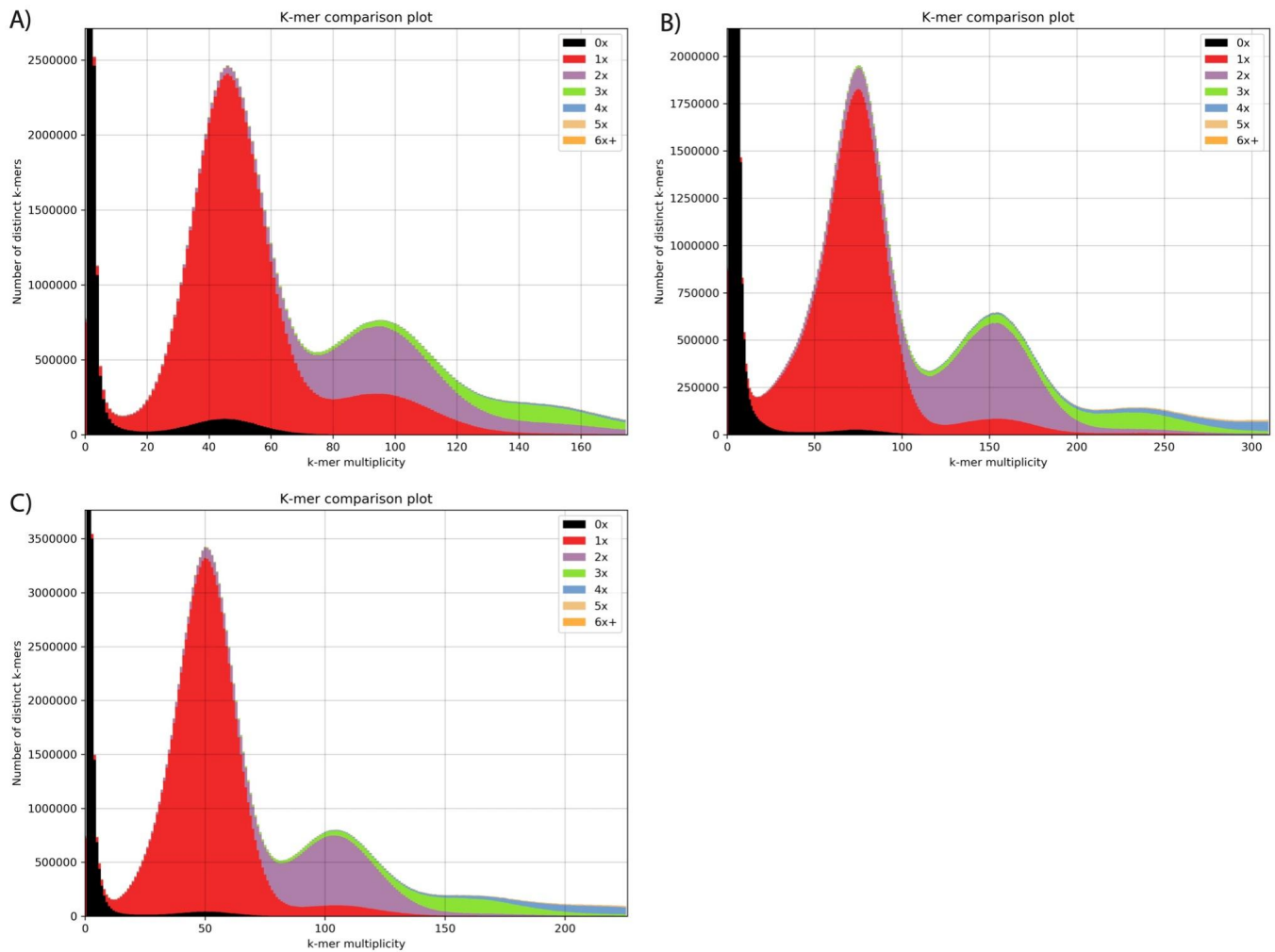
Smudgeplot <sup>1</sup> k-mer distribution analysis of the Illumina reads used for polishing the contigs predicted a triploid (3n 'AAB') genome structure for *M. incognita* (A) and tetraploid (4n 'AABB') genome structures for both *M. javanica* (B) and *M. arenaria* (C).

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



Estimation of genome size and nucleotide divergence between homoeologous genome copies using GenomeScope2 <sup>1</sup>. **(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 280.6Mb) 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

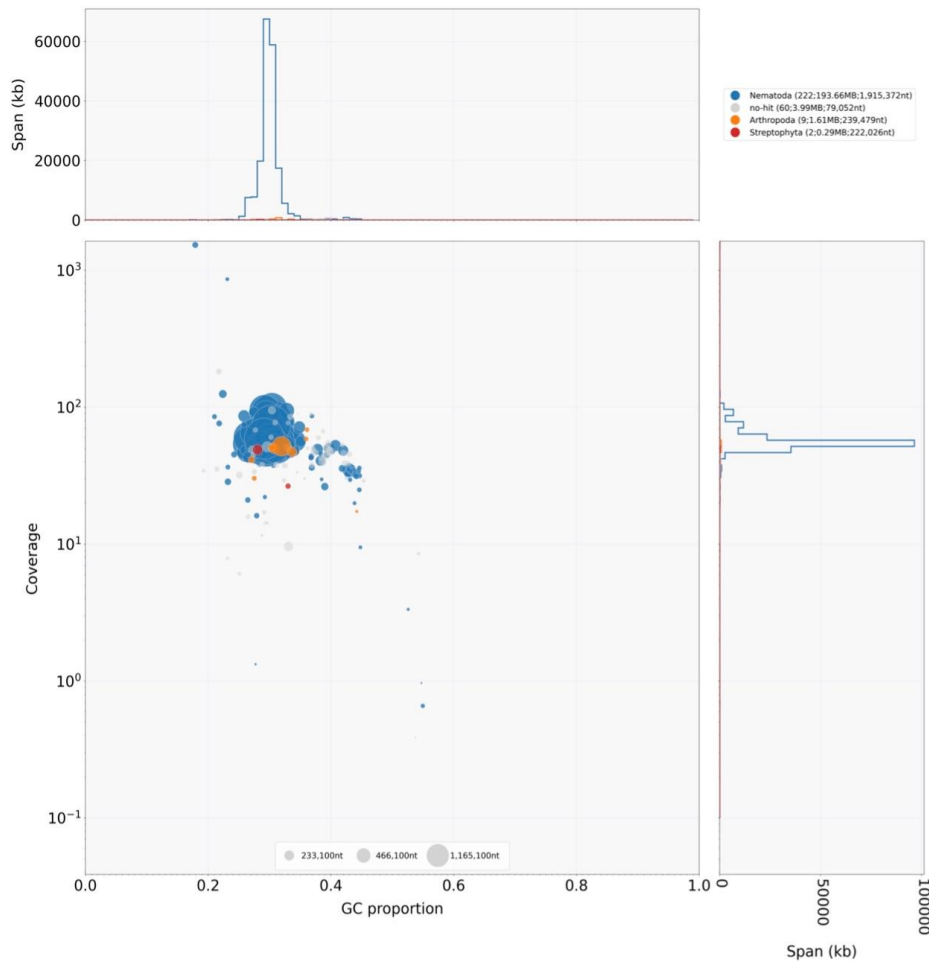


KAT<sup>2</sup> comparative analysis of the distribution of Illumina reads k-mers in the *M. incognita* (A), *M. javanica* (B) and *M. arenaria* (C) genome assemblies. Very few k-mers in the reads were not represented in the assembly (0X black curve), suggesting almost all the information present in the reads was retrieved in the assemblies. This observation and the congruence between genome assembly sizes and those estimated via flow cytometry<sup>3</sup> suggest complete genomes with homoeologous genome copies having been mostly unzipped during assembly.

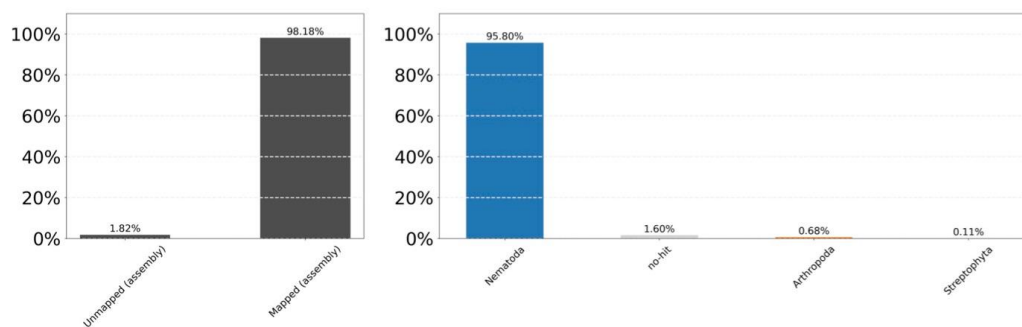
## Supplementary Figure 4: genome contamination assessment with blobtools

### 1) *M. incognita*

A)



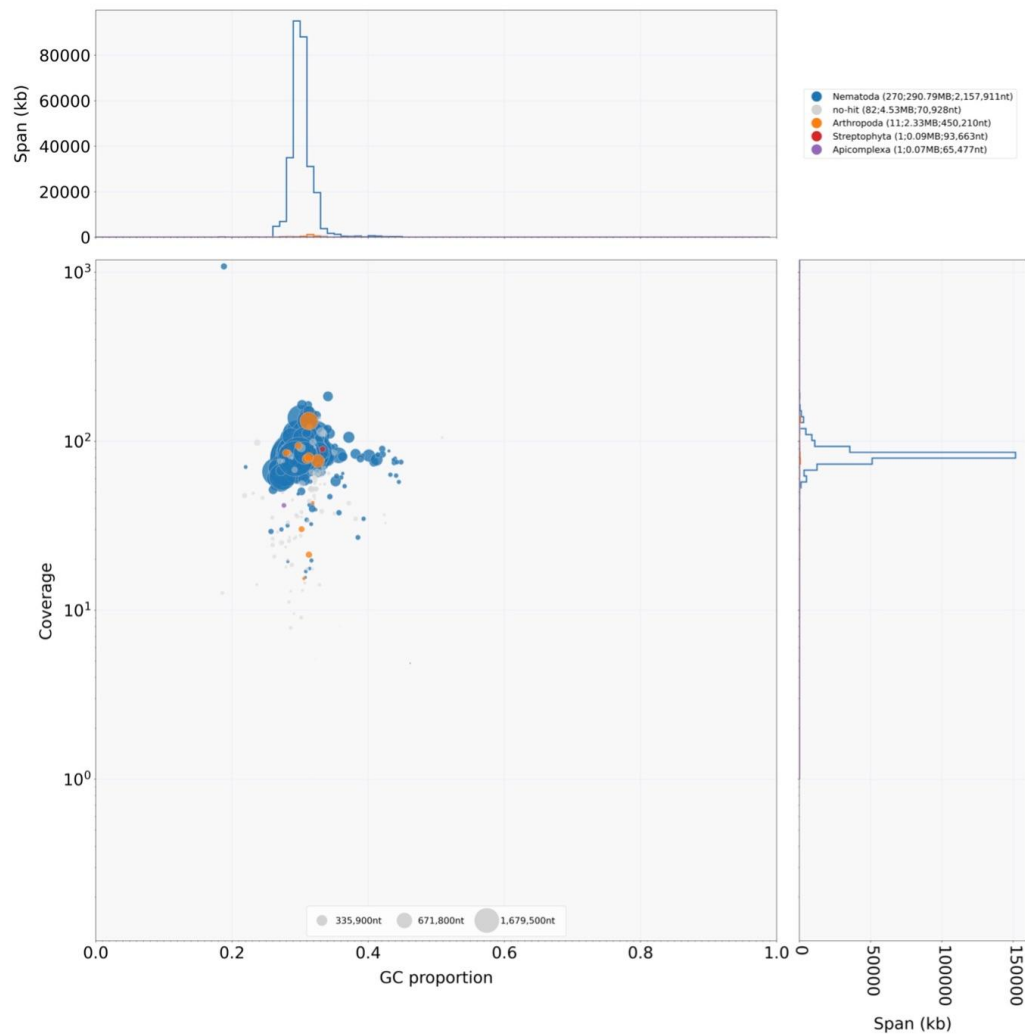
B)



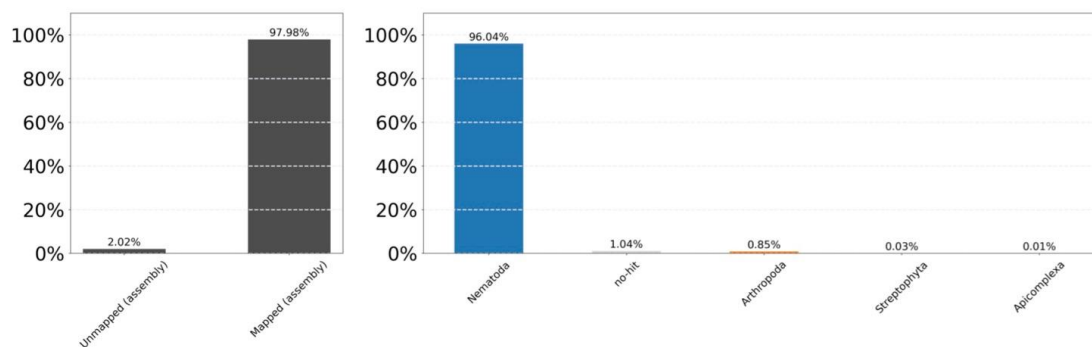
No evident trace of contamination could be identified. Two contigs with very high coverage corresponded to the mitochondrial genome of *M. incognita* and were removed from the final nuclear genome assembly.

## 2) *M. javanica*

A)



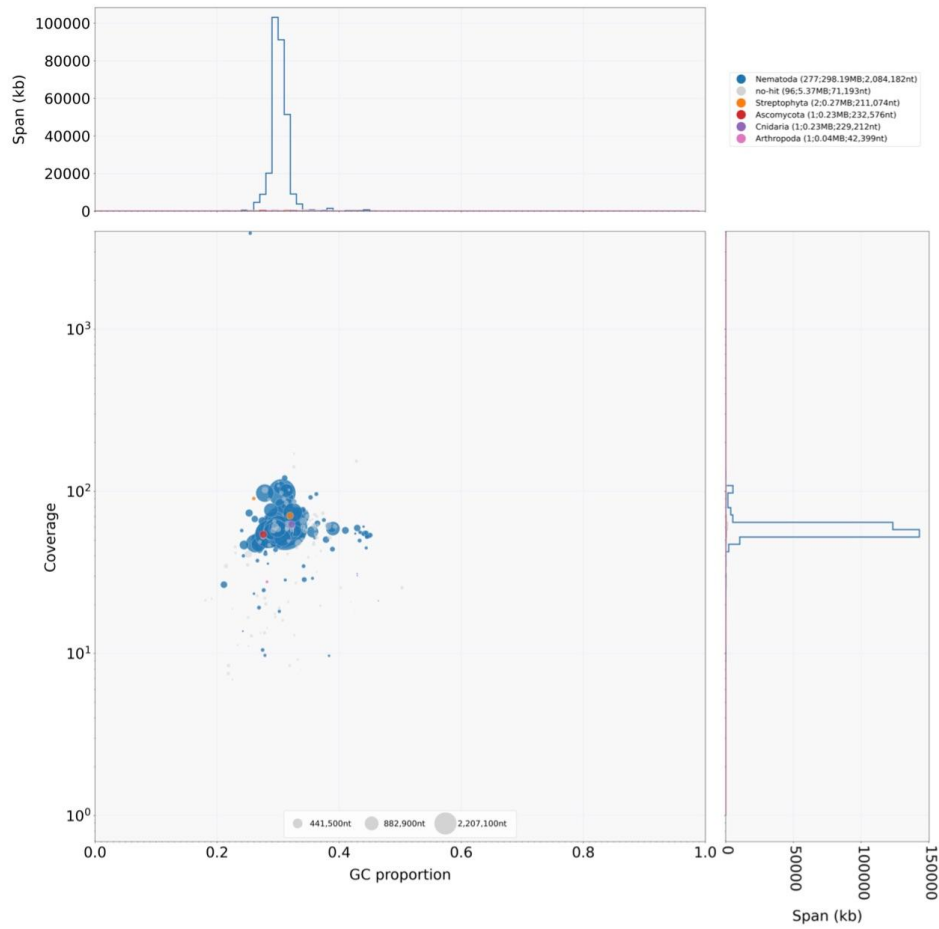
B)



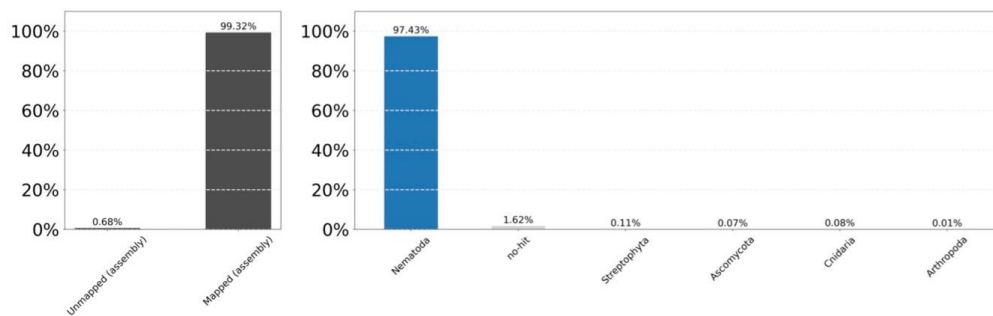
The majority of the contigs had approximately the same coverage, and one contig identified with a high coverage was annotated as the mitochondrial genome of *M. javanica* and removed from the assembly. A low percentage of the contigs was mapped as Arthropoda (0.85%), but this can be the result of lack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

### 3) *M. arenaria*

A)



B)



One contig was identified with an 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

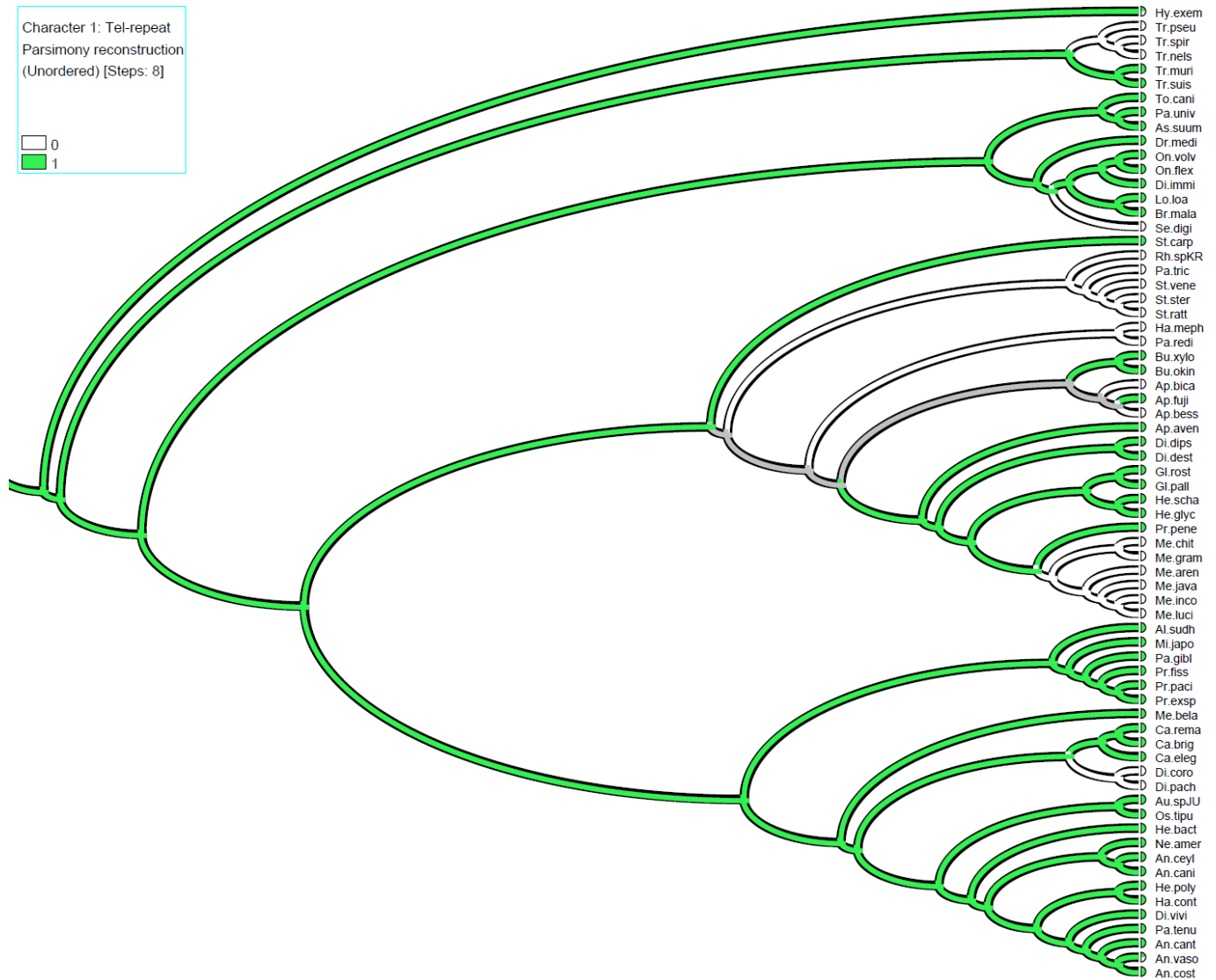
| Feature / species                             | <i>Minc</i> | <i>Mjav</i> | <i>Mare</i> | <i>Mruci</i> |
|-----------------------------------------------|-------------|-------------|-------------|--------------|
| Predicted genes                               | 54,816      | 64,689      | 63,434      | 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 <sup>1</sup>  | 15.8%       | 23.8%       | 24.3%       | 16.4%        |

1: regions masked by EuGene <sup>4</sup> using RED <sup>5</sup> prior to gene prediction because they are repetitive with no evidence for transcription.

## Supplementary Table 2: Distribution of (TTAGGC)<sub>n</sub> repeat and telomere-associated proteins across nematode genomes.

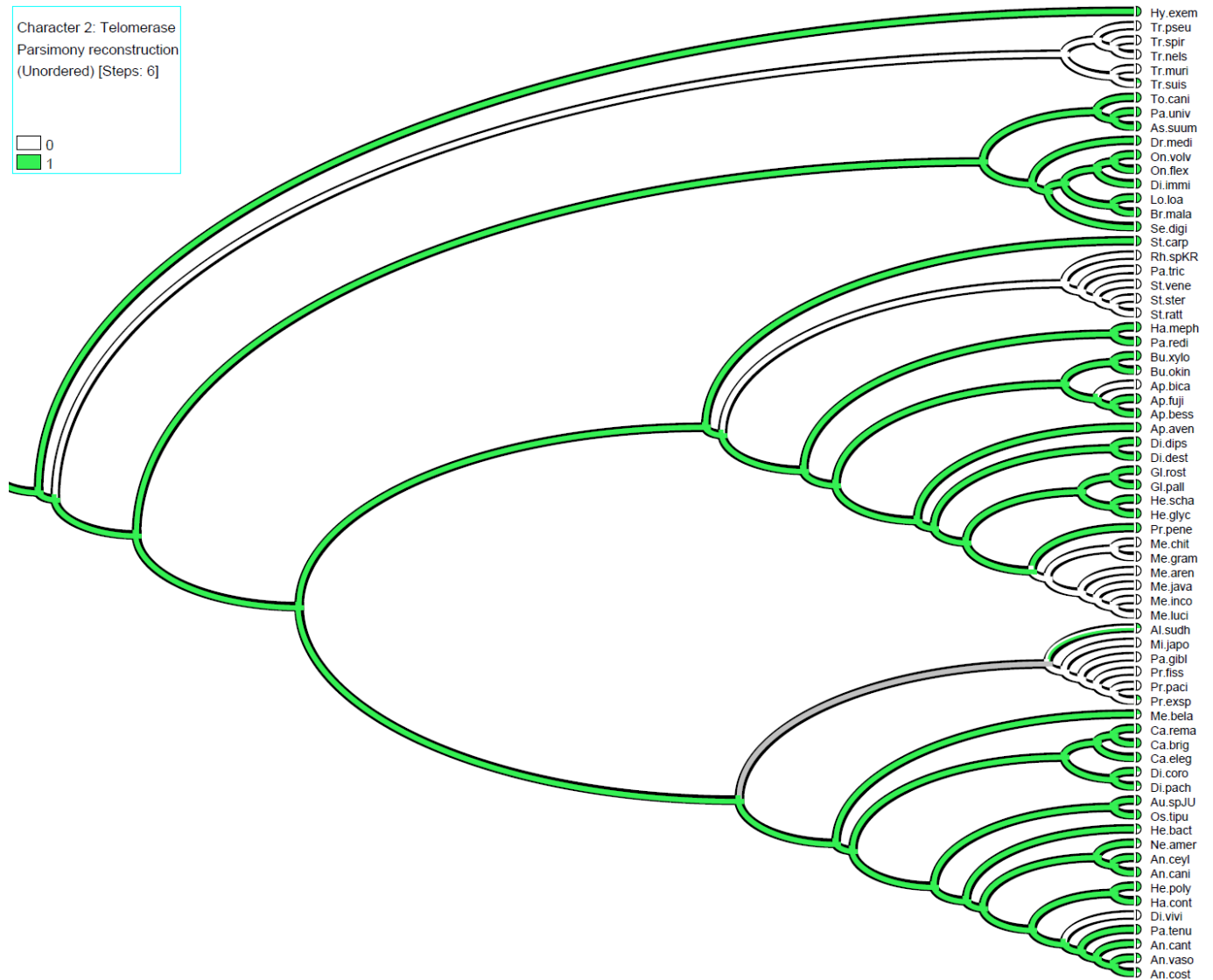
Table available online at Recherche Data Gouv: <https://doi.org/10.57745/9DGKGP>

## Supplementary Figure 5: telomeric repeat ancestral state reconstruction



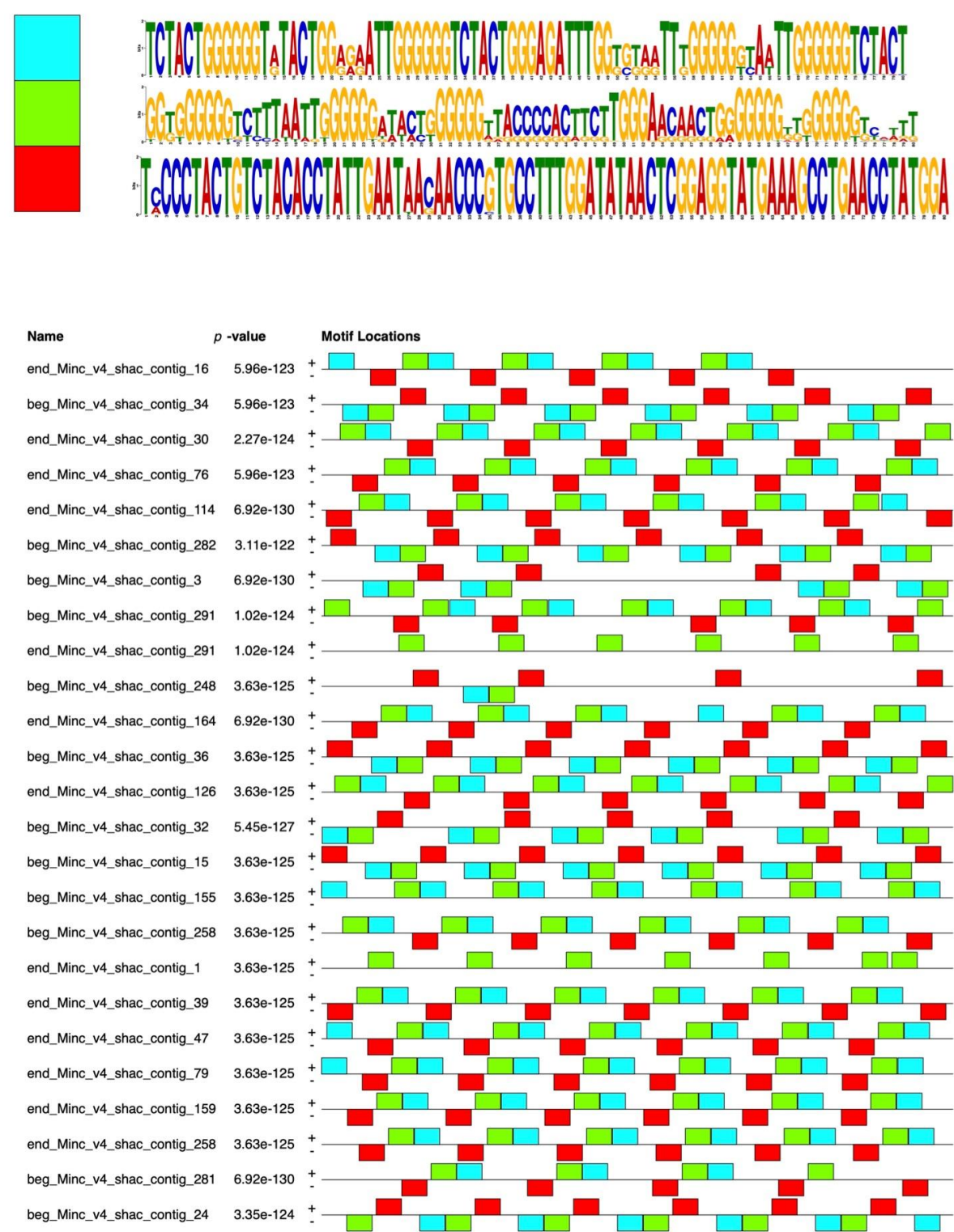
Based on presence / absence of the (TTAGGC) $n$  telomeric repeat of *C. elegans* in nematode genomes, ancestral states were reconstructed using parsimony in Mesquite <sup>6</sup>. Green means (TTAGGC) $n$  present and white absent, gray means presence and absence are equi-parsimonious. In the tardigrade *Hypsibius dujardini*, another simple repeat, (GATGGGTTTT) $n$ , was described as a candidate telomeric repeat <sup>7</sup>.

## Supplementary Figure 6: 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 Mesquite <sup>6</sup>. Green means telomerase present and white absent.

Supplementary Figure 7: enriched motif at *M. incognita* contig ends



Enriched repeated motifs by increasing e-value and their distribution on contig extremities (first and last 2kb) in the *M. incognita* genome.

## Supplementary Figure 8: consensus of the enriched motif at *M. incognita* contig ends



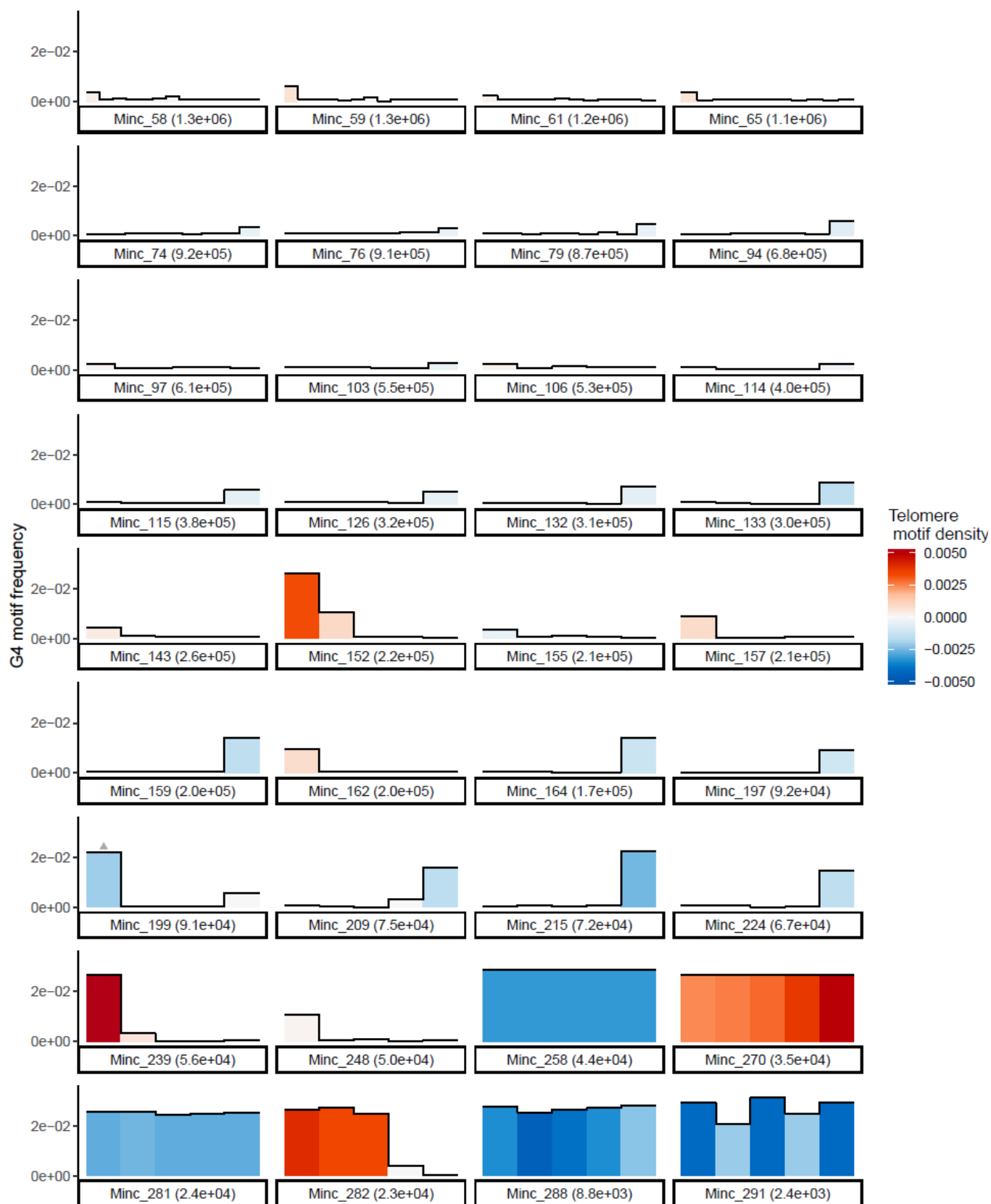
Positions of the primers used for PCR and FISH experiments are indicated as Minc\_Telomere-R and Minc\_Telomere\_L

**GGAGTGACATATTTTGTCTGGGGG**TCTACTGGGGGTCA**TTTGGGGGTCTACTGGGGGTGTACTGGAGAATT  
GGGGGTCTACTGGGAGATTGGGGGTGGGGGTCA**TTTGGGGGTCTACTTGGGTGGGGGTCTTTAATTGG**  
GGGATACTGGGGGTACCCCACTTCTTGGGAACAACCTGGGGGTGGGGGTCA**TTTGGGGGTCTAGTTTC**  
CCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTCGGAGGTAT**GAAAGCCTGAAC**  
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 9: distribution of repeat patterns and G4-quadruplex on *M. incognita* contigs



**Density of repeat patterns and G4-quadruplex along Minc contigs.** The density of repeated complex patterns 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 in the same 100kb windows. Only contigs containing repeat arrays and smaller than the N80 value are represented, the same information for the bigger repeat-containing contigs is available in (Figure 3).

### Supplementary Table 3: position of Minc repeats on Minc contigs

Table available online at Recherche Data Gouv: <https://doi.org/10.57745/YSMOP2>

### Supplementary Table 4: position of Minc repeats on Mare contigs

Table available online at Recherche Data Gouv: <https://doi.org/10.57745/VMVPWQ>

### Supplementary Table 5: position of Minc repeats on Mjav contigs

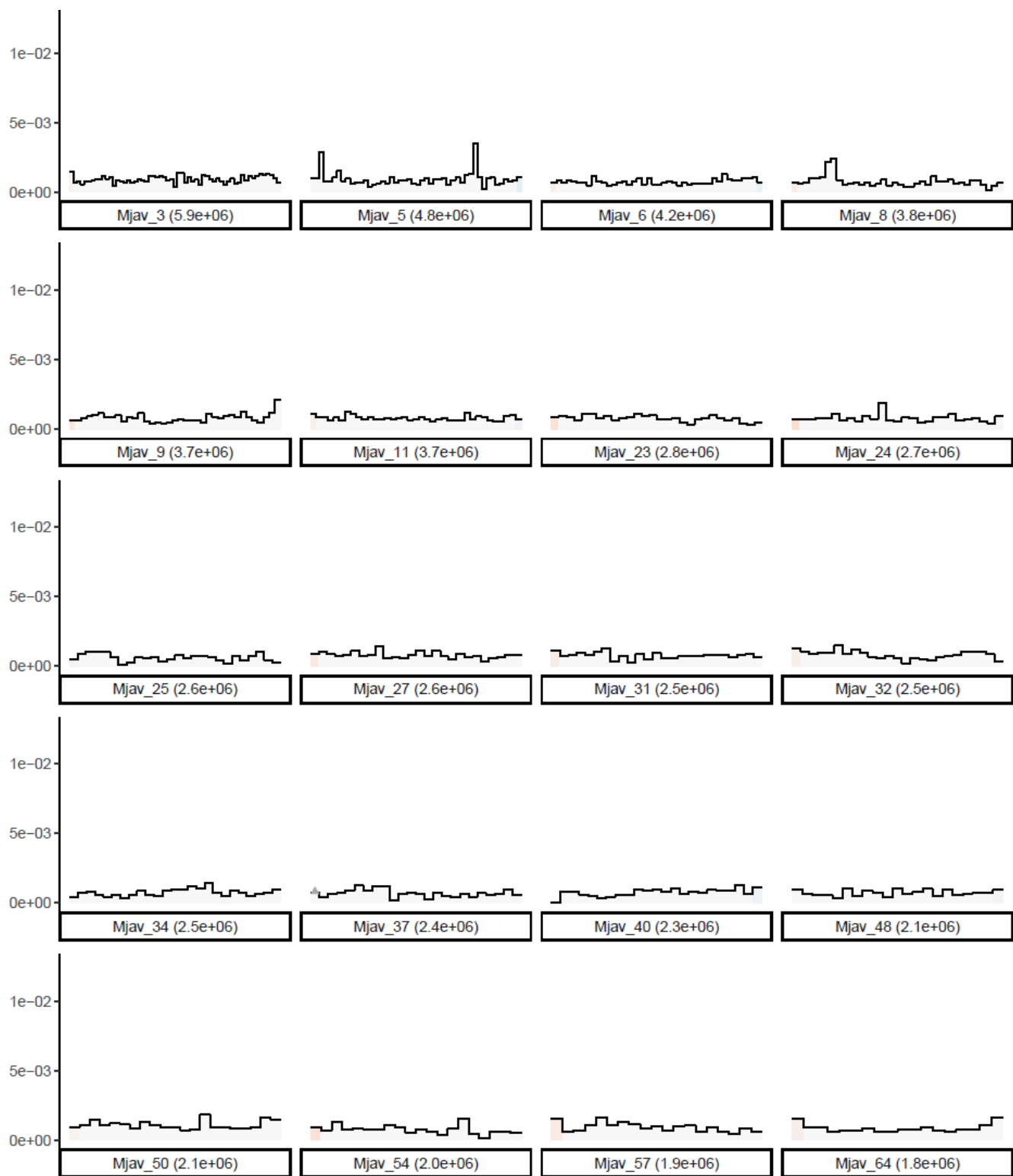
Table available online at Recherche Data Gouv: <https://doi.org/10.57745/AUWWIA>

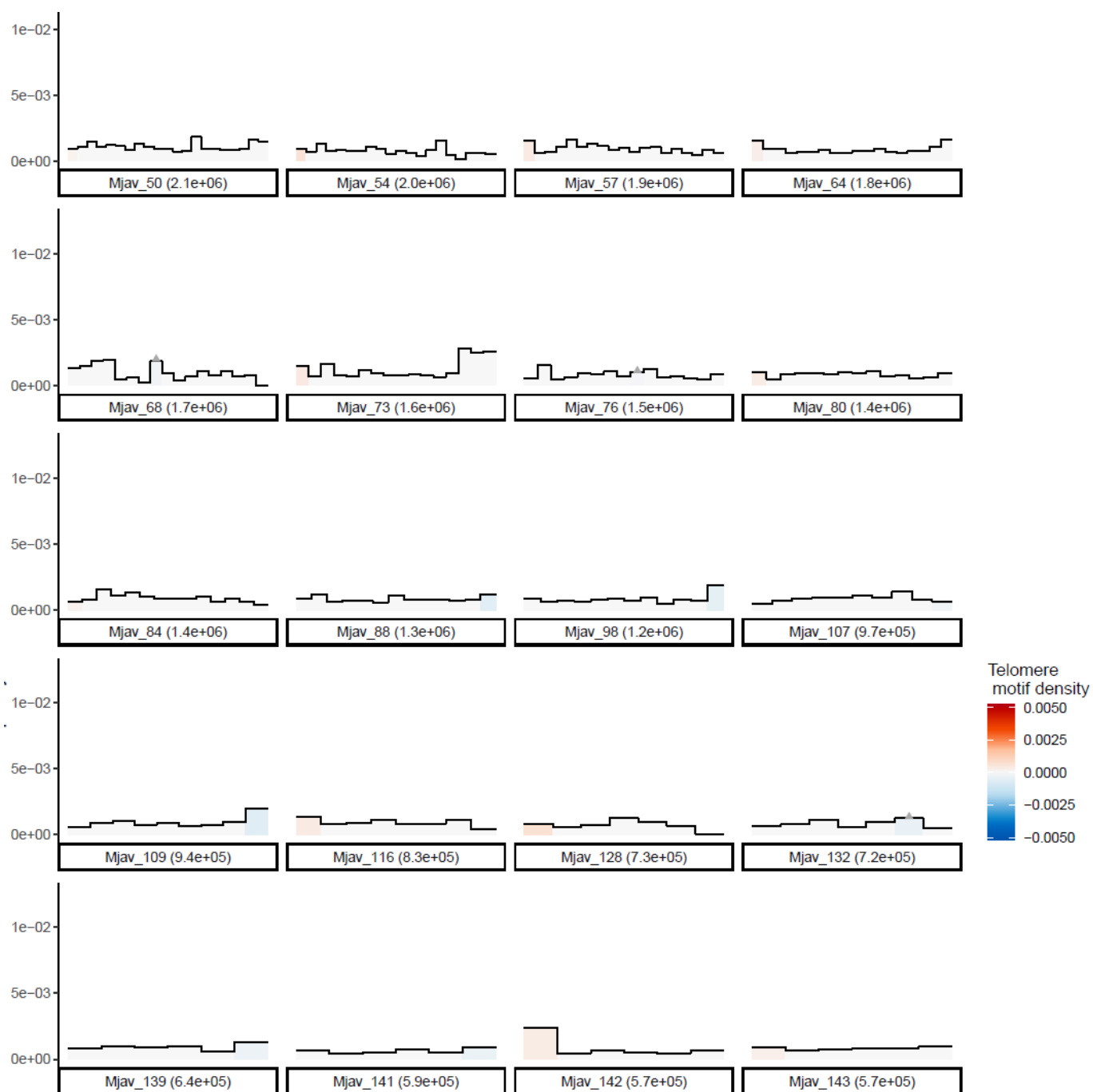
Supplementary Figure 10: distribution of repeat patterns and G4-quadruplex on *M. javanica* contigs

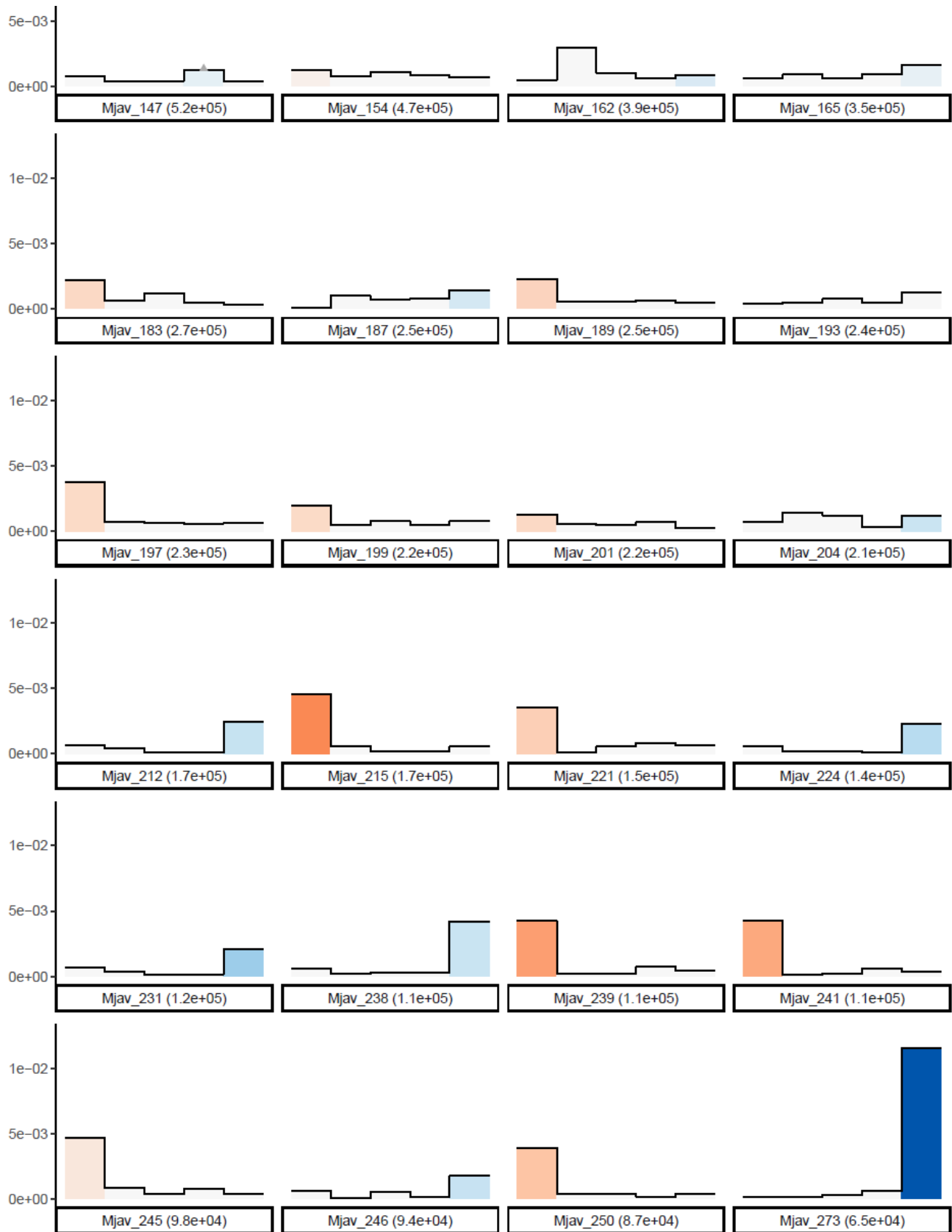
A )Consensus sequence of *M. javanica*

TnAAAAAGGGGGTCGGTCAGTTCAGTGC GGGATTTTCAGTGC GGGTTATTTTCnGTGCGGnnnGnnTAGnGnnnTnnnGAAGGTCTA  Motif 2  
 AGTGCCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTCGGAGGTATGAAAGCCTGAACCTATGGAGAGGATT  Motif 1  
 TGAGTCGGnGCAGTTGAAGTTGTAGACGTCTGGCAGGAGCAGTTGAAGTTGTAGACGGCCnGTAGAAGCnGTTGAAGTTGTAGACGGC  Motif 3

TnAAAAAGGGGGTCGGTCAGTTCAGTGC GGGATTTTCAGTGC GGGTTATTTTCnGTGCGGnnnGnnTAGn  
 GnnnTnnnGAAGGTCTAAGTGCCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTC  
 GGAGGTATGAAAGCCTGAACCTATGGAGAGGATTTGAGTCGGnGCAGTTGAAGTTGTAGACGTCTGGCAG  
 GAGCAGTTGAAGTTGTAGACGGCCnGTAGAAGCnGTTGAAGTTGTAGACGGC







**B) Density of repeat patterns and G4-quadruplex along *M. javanica* contigs.** The density of repeated complex patterns 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 in the same 100kb windows.

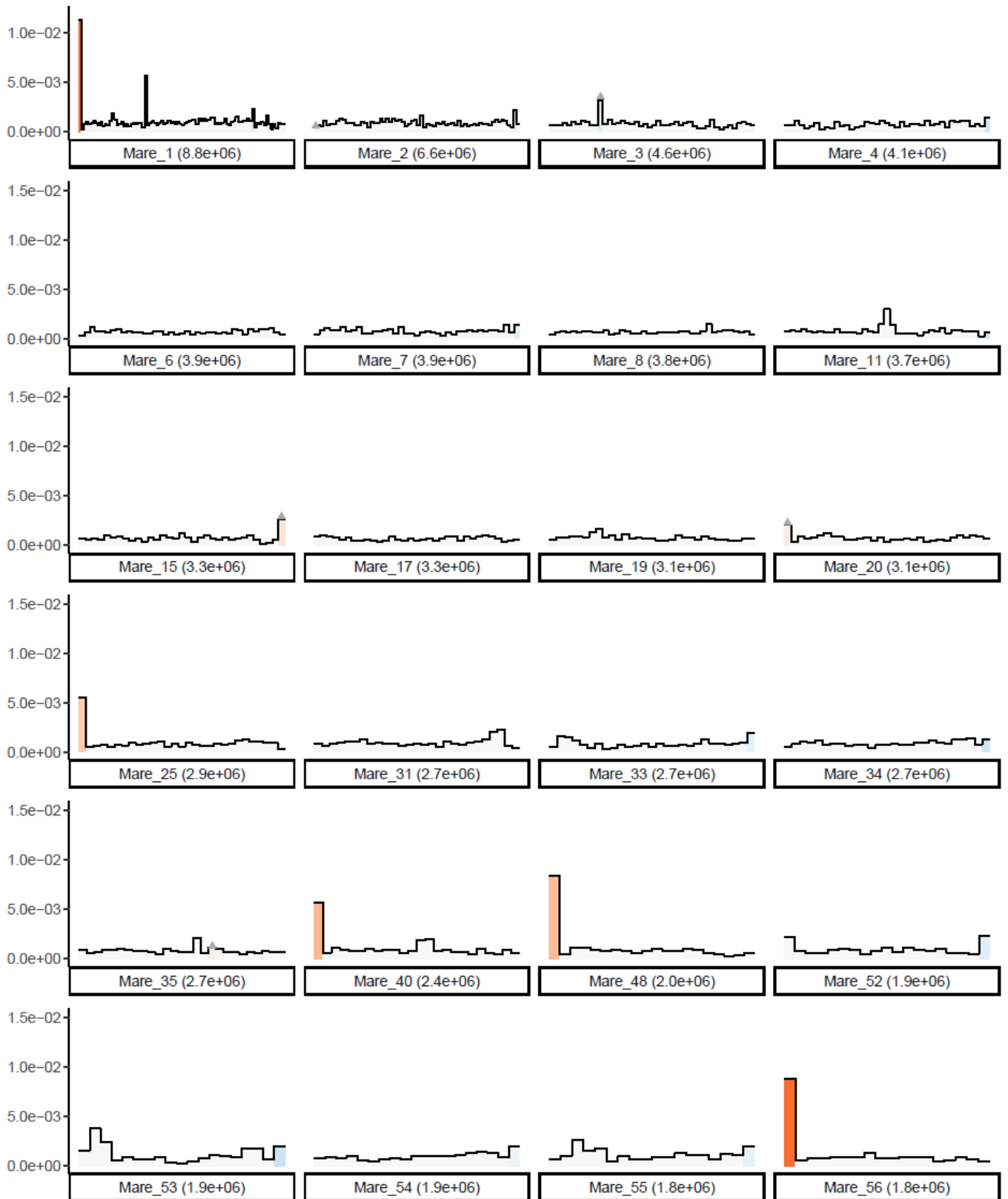
## Supplementary Figure 11: distribution of repeat patterns and G4-quadruplex on *M. arenaria* contigs

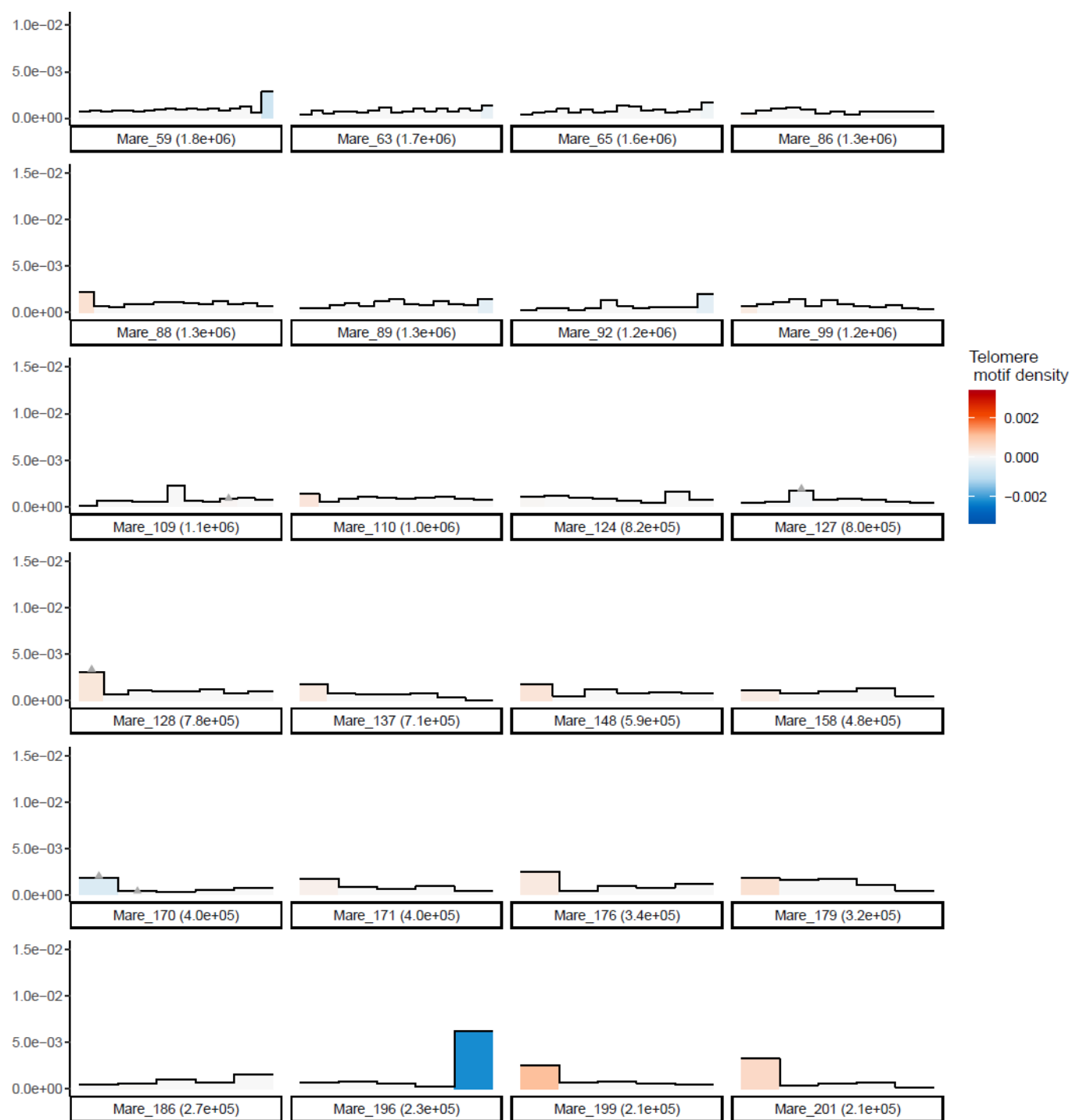
### A) Consensus sequence of *M. arenaria*

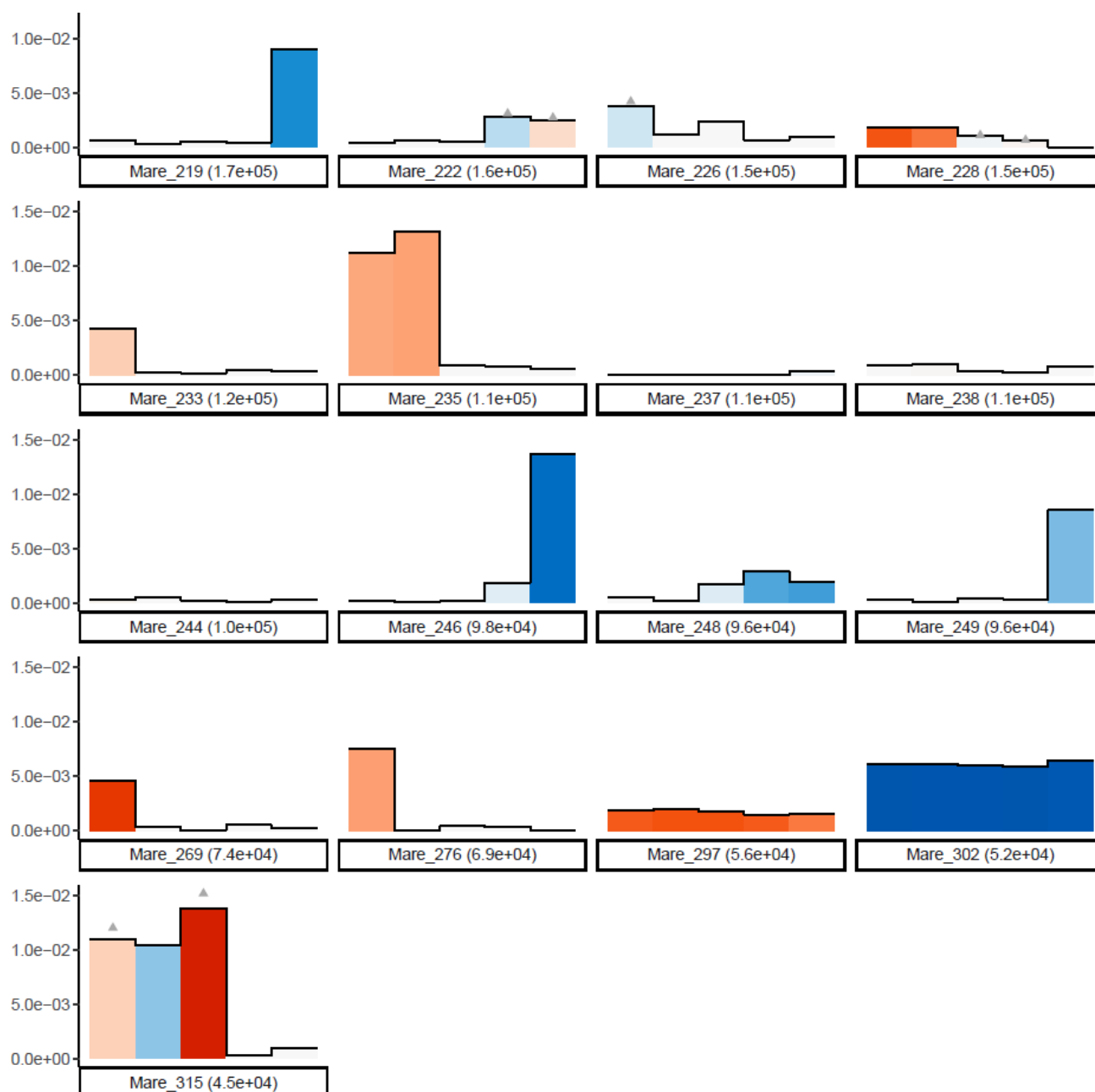
CAAATCTAAGGTCTAAGTGCCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTCGGAGGTATGAAAGCCTGAACCTA<sub>nn</sub>C  
 C<sub>n</sub>CCCCCCCCCCTTC<sub>n</sub>GCACTTCAACGG<sub>n</sub>CCGTCTACAACCTCAACTGCTCCTGCCAGACGTCTACAACCTCAACTGCTCCTGCAG<sub>n</sub>ACC<sub>n</sub>T<sub>n</sub>  
 TTGGACCCGGGGGGTCATTAGAGTACCGTGTGGGGCCGAGGGACCATTAGAGTCGGTGCAGGAGAAGTTGTAGACGTCTGGCAGGAGCTG

 Motif 2  
 Motif 3  
 Motif 1

CAAATCTAAGGTCTAAGTGCCCTACTGTCTACACCTATTGAATAACAACCCGTGCCTTTGGATATAACTCGG  
 AGGTATGAAAGCCTGAACCTA<sub>nn</sub>CC<sub>n</sub>CCCCCCCCCCTTC<sub>n</sub>GCACTTCAACGG<sub>n</sub>CCGTCTACAACCTC  
 AACTGCTCCTGCCAGACGTCTACAACCTCAACTGCTCCTGCAG<sub>n</sub>ACC<sub>n</sub>T<sub>n</sub>TTGGACCCGGGGGGTCATTAG  
 AGTACCGTGTGGGGCCGAGGGACCATTAGAGTCGGTGCAGGAGAAGTTGTAGACGTCTGGCAGGAGCT  
 G







**B) Density of repeat patterns and G4-quadruplex along *M. arenaria* contigs.** The density of repeated complex patterns 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 in the same 100kb windows.

Frequency of G4-quadruplexes along contigs with predicted telomeric sequences in *M. arenaria*.

## Supplementary Table 6: position of Minc repeats on Mluci contigs

Table available online at Recherche Data Gouv: <https://doi.org/10.57745/RLLZRM>

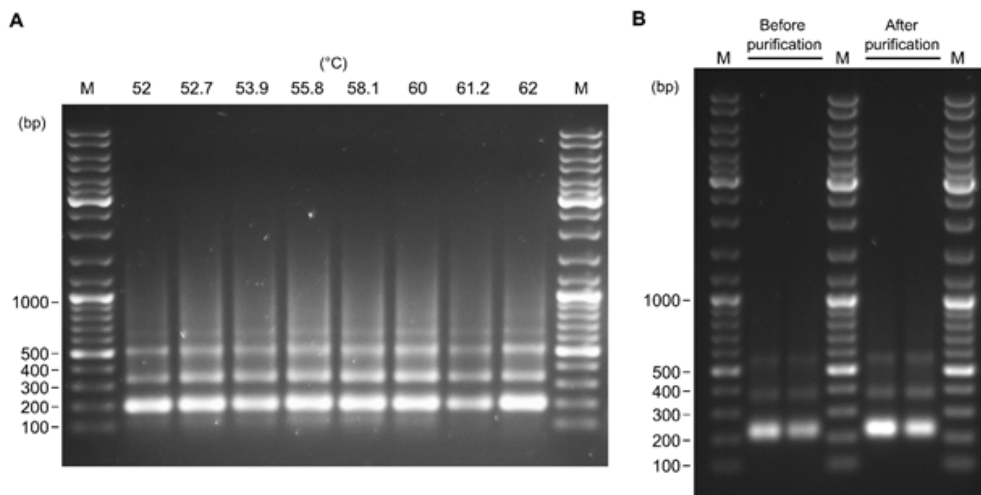
## Supplementary Table 7: annotation of (retro)transposons and other repeats

The annotation of transposable elements and other repetitive regions on the genomes of *M. incognita*, *M. arenaria* and *M. javanica* was performed using EDTA <sup>8</sup> 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 12. Gel electrophoresis of amplified telomere sequence.

A) Gradient PCR using primer pair (Supplementary Figure 12) and temperature range as listed in Materials and Methods section. B) Labeled probe for telomere sequence using biotin-dUTP nucleotide in PCR reaction.



## Supplementary Table 8: de novo assembled Minc transcripts containing the repeat and matching the genome

Table available online at Recherche Data Gouv: <https://doi.org/10.57745/XWLWCJ>

## Supplementary Table 9: ISO-seq Minc transcripts containing the repeat and matching the genome

Table available online at Recherche Data Gouv: <https://doi.org/10.57745/CHCR3F>

## Supplementary Table 10: Raw ONT sequencing library information.

Sequencing libraries were produced using the Minion or Promethion platform, with flow-cell version 9.4.1 in both cases. Libraries sequenced with a Minion device were then base-called using guppy dna\_r9.4.1\_450bps\_sup.cfg configuration file. Libraries sequenced with a Promethion device were then base-called using guppy dna\_r9.4.1\_450bps\_sup\_prom.cfg configuration file. All base calling was made in super high accuracy mode.

| Species             | Raw sequencing library  | Platform   | Accession numbers |
|---------------------|-------------------------|------------|-------------------|
| <i>M. incognita</i> | BZU_AD_ONT_1_FAK08876_A | Minion     | ERX10638972       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH89151_A | Minion     | ERX10638973       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAF18555_A | Minion     | ERX10638974       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH58676_A | Minion     | ERX10638975       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH54346_A | Minion     | ERX10638976       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH43582_A | Minion     | ERX10638977       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH54287_A | Minion     | ERX10638978       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH63986_A | Minion     | ERX10638979       |
| <i>M. incognita</i> | BZU_AA_ONT_1_FAH60611_A | Minion     | ERX10638980       |
| <i>M. incognita</i> | BZU_AA_ONT_1_PAD23691_A | Promethion | ERX10638981       |
| <i>M. incognita</i> | BZU_AA_ONT_1_PAC22366_A | Promethion | ERX10638982       |
|                     |                         |            |                   |
| <i>M. javanica</i>  | BZU_AH_ONT_1_FAK26065_A | Minion     | ERX10638983       |
| <i>M. javanica</i>  | BZU_AC_ONT_1_FAH59483_A | Minion     | ERX10638984       |
| <i>M. javanica</i>  | BZU_AC_ONT_1_FAH54228_A | Minion     | ERX10638985       |

|                    |                         |            |             |
|--------------------|-------------------------|------------|-------------|
| <i>M. javanica</i> | BZU_AC_ONT_1_FAH43433_A | Minion     | ERX10638986 |
| <i>M. javanica</i> | BZU_AC_ONT_1_FAH58655_A | Minion     | ERX10638987 |
| <i>M. javanica</i> | BZU_AC_ONT_1_FAH54419_A | Minion     | ERX10638988 |
| <i>M. javanica</i> | BZU_AG_ONT_1_PAD26893_A | Promethion | ERX10638989 |
| <i>M. javanica</i> | BZU_AC_ONT_1_PAC15472_A | Promethion | ERX10638990 |
|                    |                         |            |             |
| <i>M. arenaria</i> | BZU_AH_ONT_1_PAD26768_A | PromethION | ERX10639152 |
| <i>M. arenaria</i> | BZU_AG_ONT_1_FAK26121_A | Minion     | ERX10638991 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH54179_A | Minion     | ERX10638992 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH58521_A | Minion     | ERX10638993 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH58614_A | Minion     | ERX10638994 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH54190_A | Minion     | ERX10638995 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH43612_A | Minion     | ERX10638996 |
| <i>M. arenaria</i> | BZU_AB_ONT_1_FAH59474_A | Minion     | ERX10638997 |

Supplementary Table 11: 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       |

## 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).