Unveiling the genome assembly of the polyploid and quarantine root-knot nematode, *Meloidogyne enterolobii*

Authors

Marine Poullet[1], Hemanth Gopal[2], Corinne Rancurel[1], Marine Sallaberry[3], Celine Lopez-Roques[3], Joanna Lledo[3], Sebastian Kiewnick[2*], Etienne GJ Danchin[1*]

Affiliations

1. Institut Sophia Agrobiotech, INRAE, Université Côte d'Azur, CNRS, 400 routes des Chappes, 06903 Sophia-Antipolis, France
2. Julius Kühn-Institut, Institute for Plant Protection in Field Crops and Grassland, Messeweg 11-12, 38104 Braunschweig, Germany
3. Plateforme Génomique GeT_PlaGe, INRAE, 24 Chemin de Borde Rouge – Auzeville, 31326 Castanet – Tolosan, France

*These authors jointly supervised this work

Corresponding author(s): Marine Poullet (marine.poullet@inrae.fr); Etienne GJ Danchin (etienne.danchin@inrae.fr)

Abstract

Root-knot nematodes of the genus *Meloidogyne* are obligatory plant endoparasites that cause substantial economic losses to the agricultural production and impact the global food supply. These plant parasitic nematodes belong to the most widespread and devastating genus worldwide, yet few measures of control are available. The most efficient way to control root-knot nematodes (RKN) is deployment of resistance genes in plants. However, current resistance genes that control other *Meloidogyne* species are mostly inefficient on *M. enterolobii*. Consequently, *M. enterolobii* was listed as a European Union quarantine pest implementing regulation. To gain insight into the molecular characteristics underlying its parasitic success, exploring the genome of *M. enterolobii* is essential. Here, we report a high-quality genome assembly of *Meloidogyne enterolobii* using the high-fidelity long-read sequencing technology developed by Pacific Biosciences, combined with a gap-aware sequence transformer, DeepConsensus. The resulting genome assembly spans 273 Mbp with 556 contigs, a GC% of 0.30 ± 0.042 and an N50 value of 2.11Mb, constituting a useful platform for comparative, population and functional genomics.

Background & Summary

Root-knot nematodes (RKN) belong to the genus *Meloidogyne*, and are among the most destructive plant-parasitic nematodes[1]. Due to their extensive geographic distribution and ability to infest a wide range of host plants, they have a detrimental impact on the yield and quality of numerous economically valuable crops[2]. At present, the Meloidogyne genus comprises more than 100 described species. However, *M. arenaria, M. incognita, M. javanica and M. hapla* are considered the most widespread and damaging, RKN species[3]. In recent years, *M. enterolobii* has received increasing attention due to its unique ability to overcome several sources of resistance against RKN[2,4,5].

The species *Meloidogyne enterolobii* was originally first described as *M. incognita* from a population obtained from the Pacara Earpod Tree (*Enterolobium contortisiliquum* [Vell.] Morong) in Hainan Island, China by Yang and Eisenback (1983)[6]. Later in 1988, Rammah and Hirschmann described a new species[7], *M. mayaguensis*, sampled from eggplant (*Solanum melongena* L.) roots from Puerto Rico. However, this species was later synonymized with *M. enterolobii*, based on the same esterase phenotype and mitochondrial DNA sequence[8,9].

Meloidogyne enterolobii has an extremely high damage potential[10], surpassing many of the other root-knot nematode species studied so far[11,12]. The reports of severe damage in high-value crops have increased in the past years[13,14]. In 2009, the European Plant Protection Organization (EPPO) performed a risk analysis, which came to the conclusion that this species was recommended for regulation and placed on the EPPO A2 list in 2010. Following numerous interceptions over the years, it was concluded that *M. enterolobii* fulfilled the conditions provided in Article 3 and Section 1 of Annex I to Regulation (EU) 2016/2031 in respect of the Union territory and therefore should be listed in Part A of Annex II to Implementing Regulation (EU) 2019/2072 as Union quarantine pest[15]. However, once damage is detected, *M. enterolobii* identification is challenging due to morphological resemblances it shares with other root-knot nematode species[11,14,16,17].

In that perspective, providing high-quality nuclear and mitochondrial genomes for this species can accelerate the development of reliable molecular markers and the understanding of the biology of *M. enterolobii*. A first version of the genome of *M. enterolobii* was published in 2017 as part of a comparative genomics analysis with other root-knot nematodes[18]. The population named L30, originated from Burkina Faso and was sequenced using Illumina short reads. Consequently, the assembled genome was quite fragmented with >46,000 contigs and an N50 length <9.3kb, precluding analyses of structural variants or conserved synteny with other *Meloidogyne* species. Nevertheless, this initial genome allowed confirming that this species

was likely polyploid, similarly to other tropical parthenogenetic root-knot nematodes[19]. More recently, a *M. enterolobii* genome assembled from PacBio RS long reads and polished with Illumina short reads was published[20]. The sequenced population named Mma-II, was isolated from infected tomatoes in a Swiss organic farm[21]. With <4,500 contigs and an N50 length of 143kb, the genome assembly represented a substantial improvement compared to the only other assembly available at this time. However, recent progress in the quality and data volume of long-read sequencing technologies[22] promises even more contiguous and higher-quality genomes even for complex polyploid species and including the Meloidogyne genus[23,24]. Therefore, we used the PacBio HiFi, highly accurate long-read sequencing technology to produce a more contiguous and reliable reference genome for this quarantine plant-parasitic nematode.

Using this technology and further improvement of the quality of the reads, we assembled the genome of the *Meloidogyne enterolobii* population (E1834), originally isolated from the roots of eggplant collected in Puerto Rico, in 556 contigs with an N50 length surpassing 2Mb and a genome assembly size of 273Mb, consistent with previous flow cytometry estimation on a population from Guadeloupe island (274.7 +- 18.52 Mb)[20]. Compared to the previous long-read version of the genome, this constitutes an improvement of the N50 contiguity by more than one order of magnitude, with the number of contigs divided by almost 10.

Further quality check of our genome assembly and comparison with previous assemblies confirmed the correct species identification for population E1834 and for the isolate from Burkina Faso previously sequenced with short reads. However, our study also revealed that the Mma-II Swiss population previously sequenced with PacBio RS underwent a contamination by *M. incognita,* which over several generations in a greenhouse, completely overtook originally described *M. enterolobii* population M-ma-II. As this population was not maintained as a single egg mass line, contamination by a highly virulent and equally pathogenic *M. incognita* population remained undetected. Mis-identification among *Meloidogyne* species is not uncommon as reported populations of *M. ethiopica* in Europe were later identified as *M. luci*[25]. Consequently, the genome assembly in that publication[20] corresponded to *M. incognita* implying no long-read-based contiguous genome for *M. enterolobii* was finally available so far. This finding also motivated us to develop a methodology based on mitochondrial genomes reconstruction and relative coverage to detect contamination between closely related species which are not detectable with standard Blobtools[26] approaches based on contigs GC content and coverage. This methodology can be reused to confirm correct species identification in sequencing projects.

103 Overall, we propose a high-quality contiguous genome for *M. enterolobii* constituting a reliable
104 resource for within- and between-species comparative genomics. The contiguity of the
105 genome enables study of structural variations and conserved synteny, which will be essential
106 towards comprehensive identification of genomic variations in relation with the host range of
107 this quarantine nematode species in Europe.
108

109 # Methods

110 **Nematode collection and DNA extraction**

111 The *Meloidogyne enterolobii* population (E1834) was originally isolated from the roots of
112 eggplant collected in Puerto Rico and has been maintained since 2005 in the *Meloidogyne* sp.
113 reference collection at The Netherlands Institute for Vectors, Invasive plants and Plant health
114 (NIVIP) Wageningen, Netherlands. In 2020, this population was kindly provided by NRC, for
115 the research conducted in the framework of the project AEGONE (No . 431627824r) and has
116 been maintained at the Julius Kühn Institut (JKI) in Braunschweig, Germany in a greenhouse
117 on the resistant tomato cultivar 'Phantasia'. Nematodes used for DNA extraction were
118 obtained from single egg mass (SEM) lines. To obtain these lines, 12 single females with egg
119 masses were carefully picked from the infected roots of tomato and second stage juveniles
120 (J2) were allowed to hatch in six well plates (SARSTEDT AG & Co. KG, Nümbrecht, DE) with 5ml
121 molecular grade water per well. After one week at room temperature (20 ± 1°C) in the dark,
122 10 wells with the highest number of hatched J2s were selected for inoculation. In addition,
123 two J2s from each egg mass were collected for DNA extraction and species verification by Real-
124 time PCR[27] and SCAR species-specific markers[28,29]. For multiplication of the SEMLs, five-week-
125 old tomato seedlings from the cultivar 'Phantasia' were transplanted into 1000 ml clay pots
126 (Risa Pflanzgefässe GmbH, Germany) containing 750ml quartz sand (0.3-1mm) supplemented
127 with slow-releasing fertilizer, Osmocote (1.5g/L). Afterwards, tomatoes were inoculated with
128 J2s obtained from the respective egg masses. Tomato plants were maintained in a greenhouse
129 at 20 to 25˚C with 16h of light and 8h of darkness. Plants were watered daily and fertilized
130 once per week with Wuxal® super solution (8:8:6; N: P: K, Hauert MANNA, Nürnberg, DE). After
131 8 weeks, the galled roots were carefully washed free of sand and the eggs and juveniles (E&J)
132 were extracted with 0.7% chlorine solution[30]. The resulting E&J suspension was counted to
133 identify the line with the highest reproduction rate. The SEML number 4 was therefore
134 selected (Table 1) for further experiments and production of DNA.
135

	Number of Egg & Juveniles per root
SEML 1	43,200 ± 577.35
SEML 2	22,7800 ± 1285.82
SEML 3	56,600 ± 1604.16
SEML 4	454,800 ± 1442.22
SEML 5	230,400 ± 945.16
SEML 6	3000 ± 503.32
SEML 7	187,800 ± 2457.64
SEML 8	170,000 ± 416.33
SEML 9	109,000 ± 1222.02
SEML 10	189,800 ± 901.85

Table 1: Number of newly produced eggs and juveniles per root system of 10 single eggs mass lines of *Meloidogyne enterolobii* population (E1834). Tabulated values are the mean count of Eggs and Juveniles with standard error for different SEML.

DNA extraction

The selected SEML 4 was multiplied on tomato plants to obtain J2 for DNA extraction. Galled tomato roots were carefully washed free of sand and placed in a mist chamber to collect freshly hatched J2 after 14 days. The J2 suspension was purified by the modified centrifuge-floatation method[31] with a 45% sugar solution to reduce contaminations such as root debris, bacteria, fungal spores, etc... Afterwards, approximately 50,000- 70,000 J2 were transferred into 1.5ml Eppendorf tube and washed 3 times with molecular grade water. After freezing in liquid nitrogen, DNA was extracted from the homogenized sample using the MasterPure Complete DNA & RNA Purification Kit (Lucigen) following the manufacturer's protocol. The DNA was suspended in 10mM Tris-HCl buffer and the DNA concentration was determined with either a Qubit™ 4 fluorometer (Life Technologies, Singapore) or NanoDrop 2000 ™ spectrophotometer (Thermo Fisher Scientific, USA). The NanoDrop 2000™ was blanked using the respective elution buffer for the method. DNA concentration was measured using Qubit™ (1X dsDNA HS (High Sensitivity) Assay Kit, Invitrogen, #Q32853) and NanoDrop 2000™. Purity was measured using the 260/280 nm and 260/230 nm absorbance ratios of NanoDrop 2000™.

Genome sequencing and read processing

The long-fragment DNA libraries from the *M. enterolobii* population E1834 were constructed at the GeT-PlaGe core facility, INRAE Toulouse according to the manufacturer's instructions

159 “Preparing whole genome and metagenome libraries using SMRTbell® prep kit 3.0”. At each
160 step, DNA was quantified using the Qubit dsDNA HS Assay Kit (Life Technologies). DNA purity
161 was tested using the nanodrop (Thermofisher) and size distribution and degradation assessed
162 using the Femto pulse Genomic DNA 165 kb Kit (Agilent). Purification steps were performed
163 using AMPure PB beads (PacBio) and SMRTbell cleanup beads (PacBio). A DNA damage repair
164 step was performed using the « SMRTbell Damage Repair Kit SPV3 » (PacBio). A total of 9.4µg
165 of DNA was purified and then sheared at 20 kb using the Megaruptor system (Diagenode).
166 Using SMRTbell® prep kit 3.0, a Single strand overhangs removal, a DNA and END damage
167 repair step were performed on 8.3µg of sample. Subsequently, blunt hairpin adapters were
168 ligated to the library and a nuclease treatment was performed using the nuclease mix of
169 "SMRTbell® prep kit 3.0". A size selection step using a 6kb cutoff was performed on the
170 BluePippin Size Selection system (Sage Science) with “0.75% DF Marker S1 6-10 kb vs3
171 Improved Recovery" protocol. Using Binding kit 2.2 kit and sequencing kit 2.0, the primer V5
172 annealed and polymerase 2.2 bounded library was sequenced by diffusion loading with the
173 adaptive loading method onto 1 on Sequel II instrument at 90pM with a 2 hours pre-extension
174 and a 30 hours movie.
175 The Sequel II sequencing system outputs 1To of raw data into a subread file. This contains
176 unaligned base calls from high-quality regions, the complete set of base quality values and
177 kinetic measurements from the sequencing instrument. This subread file is used as input for
178 the Circular Consensus Sequencing (CCS v6.4.0) analysis to generate a draft consensus
179 sequence. Very low-quality reads (<Q9) were filtered out by using the parameter --min-
180 rq=0.88. To further improve the quality of the PacBio Sequel II reads, we have used a gap-
181 aware sequence transformer, DeepConsensus[32] (v1.1.0). As a final step, the previous subreads
182 were aligned to the draft consensus sequence using ACTC[33] with default parameters (v0.2.0)
183 and used as input to the DeepConsensus transformer-encoder. The Phred-scale read accuracy
184 score (Qconcordance) has been calculated according to Baid et al.[32] where Qconcordance = -
185 10*log10 (1-identity) and identity = matches / (matches + mismatches + deletions +
186 insertions).
187
188 **Ploidy, heterozygosity, and genome size estimation**
189 To infer the ploidy level of the *M. enterolobii* population E1834, a k-mer-based approach was
190 employed to profile the genome. The k-mer frequencies in DeepConsensus sequencing reads
191 were analyzed using KMC[34] (v3.0.0, kmc -k21 -m100 -ci1 -cs10000). In accordance with the
192 author's recommendations, canonical 21-mers were extracted using a hash and organized in
193 a histogram file using the kmc_tools transform option. To determine the appropriate coverage

194 thresholds required for the inference, the KMC histogram file is utilized as input for the cutoff
195 option in Smudgeplot[35] (v0.2.4). Subsequently, we generated a smudge plot using the
196 coverage of the identified k-mer pairs to determine ploidy.
197 To estimate genome size and heterozygosity prior to assembly, we used Genomescope[35] (v2.0)
198 on the histogram file generated from Jellyfish[36] (v2.3.0, jellyfish histo -h 1000000) and the
199 coverage thresholds produced by the Smudgeplot cutoff tool. The final genome size is
200 therefore obtained by multiplying the haploid genome size by the previously estimated ploidy
201 level.
202
203 **Genome assembly, estimation of completeness and contamination**
204 DeepConsensus reads were trimmed of remaining adapters using HifiAdapterFilt[37] (v2.0.0)
205 with default parameters. The trimmed reads were then used as input to the Peregrine-2021
206 assembler[38,39] (v0.4.11) while increasing the default number of best overlaps for each initial
207 graph (parameter --bestn 8). This parameter is optimized for highly heterozygous genomes.
208 To further assess genome assembly completeness in a reference-free approach, we used
209 Merqury's algorithm[40] (v1.3). This tool uses k-mer frequencies to evaluate a genome's base
210 accuracy and completeness. This is achieved by counting and comparing the distribution of
211 canonical 21-mers found in the assembled genome with those detected in the high-accuracy
212 DeepConsensus read set. Merqury's k-mer analysis will therefore indicate whether the
213 genome assembly has captured all the information present in the HiFi reads.
214 The screening of the contig assembly for potential contaminants by non-nematode sequences
215 was done with the Blobtools[26] pipeline (v3.2.6). DeepConsensus polished long-reads were
216 aligned to the contigs with Minimap2[41] (v2.24) and the map-hifi parameter. Each contig was
217 then assigned to a taxonomic group based on the BLAST[42] (v2.13.0+) analysis results against
218 the NCBI nucleotide (nt) database[43]. Particular attention was paid to contigs of non-nematode
219 taxa or contigs with a GC percentage deviating from the average GC content (around 30%[18])
220 of the *M. enterolobii* population E1834 to detect possible contamination. A total of 39 contigs
221 spanning ~2.5 Mbp were discarded from the assembly. The resulting assembly of 556 contigs
222 is used for downstream analyses.
223
224 **Mitochondrion assembly and functional annotation**
225 The circular mitochondrial genome sequence was reconstituted using the ALADIN[44] package
226 (v1.1) and DeepConsensus HiFi reads in input with default parameters. We employed as a
227 reference seed sequence the complete mitochondrion of *M. enterolobii* previously
228 downloaded from the GenBank database (BioProject: PRJNA927338[45]). The annotation was

229 carried out using GeSeq[46], encompassing both the tRNA, the rRNA and the protein-coding
230 genes. We set the minimum threshold of 85% for the protein and non-coding DNA search
231 identity, and we used seven Meloidogyne mitochondrial genomes as third-party references
232 (Table 2). The tRNAs prediction was also performed using third-party predictors, such as
233 tRNAscan-SE[47] (v2.0.7), ARAGORN[48] (v1.2.38), and ARWEN[49] (v1.2.3), with codon usage
234 corresponding to Metazoan and Invertebrate Mitochondrial.
235
236 **Gene prediction and genome structure determination**
237 Gene models prediction was done with the fully automated pipeline EuGene-EP[50] (v1.6.5).
238 EuGene has been configured to integrate similarities with known proteins of *Caenorhabditis*
239 *elegans* (PRJNA13758) from WormBase Parasite[51] and "nematoda" section of
240 UniProtKB/Swiss-Prot library[52], with the prior exclusion of proteins that were similar to those
241 present in RepBase[53]. The dataset of *Meloidogyne enterolobii* transcribed sequences[20] was
242 aligned on the genome and used by EuGene as transcription evidence. Only the alignments of
243 datasets on the genome spanning 30% of the transcript length with at least 97% identity were
244 retained. The EuGene default configuration was edited to set the "preserve" parameter to 1
245 for all datasets, the "gmap_intron_filter" parameter to 1 and the minimum intron length to
246 35 bp. Finally, the Nematodes-specific Weight Array Method matrices were used to score the
247 splice sites (available at this URL: http://eugene.toulouse.inra.fr/Downloads/WAM_
248 nematodes_20171017.tar.gz).

249 Genome structure analysis was conducted using MCScanX[54], with default settings. First, the
250 whole proteome of the *M. enterolobii* population E1834*,* predicted by EuGene, was self-
251 blasted with an E-value cutoff of 1e-25, a maximum of 5 aligned sequences, and maximum 1
252 high-scoring pair (hsp). Subsequently, we used gene location information extracted from the
253 GFF3 annotation file of EuGene, along with homology information based on the all-versus-all
254 BLASTP analysis, to identify and categorize each duplicated protein-coding gene into one of
255 five groups using the duplicate_gene_classifier program implemented in the MCScanX
256 package. These groups are: singleton, proximal, tandem, whole-genome or segmental
257 duplications (WGD), and dispersed duplications. Singleton refers to cases where no duplicates
258 are found in the assembly. Proximal duplicates refer to gene duplications that are on the same
259 contig and separated by 1 to 10 genes. Tandem duplicates, on the other hand, are consecutive.
260 WGD are identified when they form collinear blocks with other pairs of duplicated genes.
261 Finally, dispersed duplicates are those that cannot be assigned to any of the above-mentioned
262 categories.

263 **Species verification and validation**

264 In the following step, we further screened the genomic reads for potential contamination, this
265 time by other root-knot nematode sequences. Blobtools allows the identification of potential
266 contamination in genome assemblies, but only at distant taxonomic levels between different
267 phyla (e.g., Chordata, Nematoda, Arthropoda, …). Therefore, although contamination can be
268 detected and cleaned at this level, it remains undetectable at the intra-phylum level (e.g.
269 within Nematoda). To allow the detection of contamination by other closely related
270 nematodes at the reads level, we adapted the Blobtools pipeline to work with mitochondrial
271 genomes. The polished long-reads were aligned against complete mitochondrial sequences
272 for seven *Meloidogyne* species downloaded from the NCBI database (Table 2), using the same
273 procedure as above. Since Blobtools only works at the phylum and not species rank, we used
274 a script to create an additional hits file and assign a custom NCBI phylum TaxID to each species.
275 The seven *Meloidogyne* samples have been then temporarily assigned to a different phylum
276 for the BlobPlot visualization only (Table 2).

277

Species	Length (bp)	GenBank accession	TaxIDs	Custom TaxIDs
M. graminicola[55]	19589	NC_056772	189291	4890(Ascomycota)
M. arenaria[56]	17580	NC_026554	6304	6340 (Annelida)
M. enterolobii[45]	17053	NC_026555	390850	390850
M. javanica[57]	18291	NC_026556	6303	10190 (Rotifera)
M. incognita[58]	17662	NC_02409	6306	6656 (Arthropoda)
M. chitwoodi[58]	18201	KJ476150	59747	6447 (Mollusca)
M. oryzae[59]	17066	MK507908	325757	7711 (Chordata)

278

279 Table 2: Mitochondrial sequence data statistics for different nematodes and their
280 correspondence with the modified Blobtools analyses. Seven Meloidogyne mitochondrial
281 sequences have been analyzed for this study. For each species, a specific custom TaxID
282 corresponding to a different phylum has been used.

283

284 Species-specific SCAR (sequence characterized amplified region) markers are routinely used to
285 confirm species identity in plant-parasitic nematodes[60]. SCAR markers are locus-specific
286 fragments of DNA that are amplified by PCR using specific 15-30bp primers. In this study, we
287 retrieved primer sequences of species-specific SCAR markers from the literature
288 (Supplementary Table 1) for four Meloidogyne species with genome assemblies publicly
289 available and belonging to the same clade (*M. arenaria*, *M. incognita*, *M. javanica* and *M.*
290 *enterolobii*). We aligned all the primers to all the above-mentioned genomes with BLAST and

291 when the primer pairs matched on the same contig, we retrieved from the genome the 'virtual'
292 PCR products. After verification of consistency with the lengths from the literature, the virtual
293 PCR products were then aligned to the two previous and present versions of *M. enterolobii*
294 genome assemblies with an E-value threshold of 1e-25.
295

296 ## Data Records

297 All the PacBio HiFi sequence data as well as the genome assemblies and gene predictions
298 supporting the results of this paper have been deposited and are publicly available at the
299 EMBL-EBI's European Nucleotide Archive (ENA) under accession number PRJEB69523
300 (https://www.ebi.ac.uk/ena/browser/view/PRJEB69523)[61]. All the processed data, including
301 genome assemblies[62], gene predictions[63], and all the structural annotation[64] results have been
302 deposited and are publicly available at the Recherche Data Gouv institutional collection.
303

304 ## Technical Validation

305 **Assessing read accuracy**

306 After implementing the DeepConsensus[32] sequence transformer procedure, statistical analysis
307 showed an increase in the number of high-quality reads obtained (Fig. 1, Table 3) with long
308 fragment DNA reads of up to 26kb in length. The average length of the reads is around 11kb
309 with a total number of 2.4M reads, and a higher average Phred-scale read accuracy score
310 (Qconcordance), which increased from 31.95 before to 34 after DeepConsensus. This
311 transformer has elevated the PacBio HiFi read yield to a minimum Q30 by 10% and a minimum
312 Q40 score (99.99% read accuracy) by 70%. Furthermore, we have retrieved 198,880 long-reads
313 that were initially dismissed prior to treatment in the filter, providing us with more chances to
314 comprehensively capture the entire genome of *M. enterolobii*.
315

	Before DeepConsensus	After DeepConsensus
Longest read	26 302 bp	26 302 bp
Mean length	11 194 bp	11 187 bp
Number of reads	2 250 199	2 449 079
Number of bases	25 442 730 700	27 277 356 063
Average Qscore	31.95	34.04

316

317 Table 3: Read statistics before and after the use of DeepConsensus sequence transformer.
318 After DeepConsensus treatment, a higher number of reads with higher quality have been
319 retrieved.

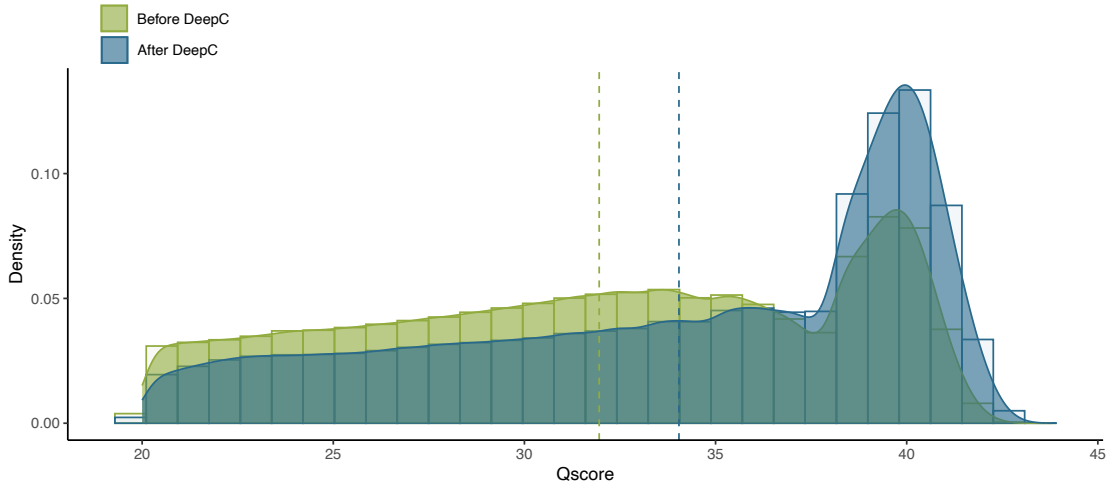


Figure 1. Distribution of raw PacBio HiFi reads before and after DeepConsensus treatment. **A**. Comparison of the Concordance Qscore before and after DeepConsensus. The average phred-scale read accuracy score has increased by two points after treatment.

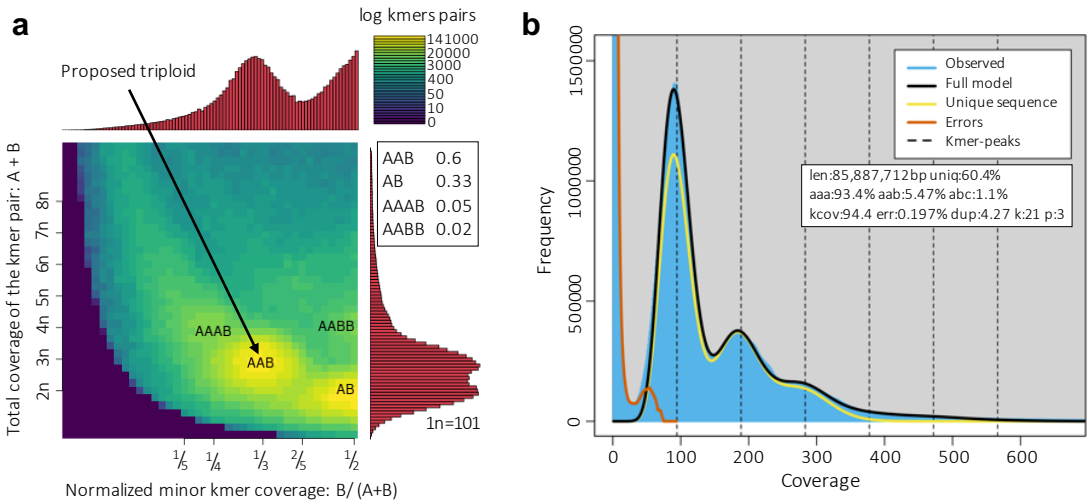
Profiling genome ploidy level, heterozygosity, and size.

Prior to assembling a genome, it is crucial to evaluate its ploidy and size. The distribution of k-mer frequencies within the DeepConsensus sequencing reads allows estimating major genome features such as ploidy level, genome size, and heterozygosity rate. As GenomeScope2[35] can only precisely examine organisms when a definite ploidy is known, we utilized first, the results of Smudgeplot[35] (Fig. 2a) to provide GenomeScope2 with estimated ploidy level. Each smudge on the graph appears to be distinct, indicating sufficient sequencing coverage for further analysis. The most prevalent smudge corresponds to a predicted triploid AAB genome for the *M. enterolobii* population E1834 (Table 4). This result is consistent with previous k-mer analysis performed on the short-reads for the L30 population from Burkina Faso[18,65].

Subsequently, we estimated the genome size using GenomeScope2 with a ploidy level of 3 (Fig. 2b). The genome size was determined by multiplying the estimated haploid genome length (85,887,712 bp) by the previously estimated ploidy level (p=3), providing an estimated genome size of 257.66 Mb.

Furthermore, the GenomeScope2 k-mer histogram of this polyploid population displays a distinct multimodal profile, with a substantial first peak located at roughly 95X, a smaller second peak at about 187X, and finally, an additional peak at 282X, typical for triploid genomes. Finally, GenomeScope2 estimated a highly heterozygous genome (6.5 % estimated on average), consistent with a previous estimation of ca. 6.1% on the L30 population from Burkina Faso[18]. It should be noted that the term heterozygosity does not exactly apply here as

346 we do not measure divergence between homologous chromosomes in a diploid genome but
347 between the AAB subgenomes in a triploid species. Therefore, we will refer to average
348 nucleotide divergence between subgenomes in the rest of the manuscript.
349



350

351 **Figure 2**. Genome profiling of *M. enterolobii*. **A**. Smudgeplot of *M. enterolobii* extracting 21-
352 mers from DeepConsensus reads. The color intensity of each smudge reflects the approximate
353 number of k-mers per bin. *M. enterolobii* E1834 population is proposed as a triploid organism.
354 **B**. GenomeScope2 k-mer profile and estimated parameters for the triploid nematode *M.*
355 *enterolobii*. Coverage (kcov), error rate (err.), haploid genome size estimation (len.), k-mer size
356 (k) and ploidy level (p). The peak heights are proportional to the species' heterozygosity. *M.*
357 *enterolobii* shows a high heterozygosity.
358

Peak	Kmers #	Kmers proportion	Summit B/ (A+B)	Summit A+B
AAB	5 952 405	0.60	0.34	309.70
AB	3 243 405	0.33	0.49	191.27
AAAB	508 560	0.05	0.24	451.82
AABB	213 490	0.02	0.49	404.45

359
360 Table 4. Summary of peaks detected by Smudgeplot. This result proposes that the *M.*
361 *enterolobii* E1834 population is triploid.
362
363 ***De novo* genome assembly**
364 After filtering and elimination of the contaminated and mitochondrial contigs, the resulting
365 genome of the *Meloidogyne enterolobii* population E1834 is assembled in 556 contigs with a

366 total size of 273 Mbp. The corresponding contig N50 length is equal to 2.11Mb, with the
367 longest being 8.3 Mb, long.
368 The genome assembly size is congruent with the total DNA content estimated by flow
369 cytometry (274.69 $\pm$ 18.52 Mb) on another population of *M. enterolobii* from Guadeloupe
370 island[20], indicating that the assembly represents a complete *M. enterolobii* genome.
371 However, the genome size estimated by analysis of k-mer distribution (257.66 Mb) is lower
372 than the assembly size and in the lower range of the flow cytometry evaluation. A previous
373 study showed that the accuracy of genome size estimation based on k-mer frequencies can be
374 affected by repeats, high heterozygosity and sequencing errors[66]. This suggests that the high
375 heterozygosity rate or repeat-richness in the *M. enterolobii* genome could have played a role
376 in this underestimation.
377 To further assess genome assembly quality metrics and evaluate genome's base accuracy and
378 completeness we used Merqury[40]. In the Merqury spectrum produced (Fig. 3a), the first and
379 prominent 1-copy peak at a ~100X multiplicity corresponds to k-mers in the reads detected
380 only one time in the assembly. This can be interpreted as heterozygous regions between the
381 three subgenomes. The second peak at twice this multiplicity (~200X) corresponds to
382 homozygous k-mers present in the reads and detected two times in the assembly. This most
383 likely represents regions identical between two of the three AAB subgenomes. Similarly, most
384 of the k-mers detected at 3 times in the assembly, probably represent regions identical
385 between the three AAB subgenomes. Conversely, the grey peak at low multiplicity represents
386 rare k-mers which solely exist within the read set and are probably due to sequencing errors.
387 This Merqury plot indicates a lack of missing content as there is no subsequent grey peak at
388 the 1-copy peak (~100X coverage). Additionally, the AAB subgenomes are divergent enough
389 to have been mostly separated (or unzipped) during the assembly, as there is no smaller
390 second grey peak beneath the 2-copy peak at twice the coverage (~200X). This can also be
391 observed in the coverage plot provided by bedtools[67] v2.29.0 (Fig. 3b), where the coverage
392 depth for each base on each contig has been computed. We can clearly see a prominent peak
393 located at roughly 101X corresponding to the haploid coverage found in the k-mers with
394 Smugeplot (Fig. 2a), and a shoulder at roughly twice the haploid coverage probably
395 represented few identical regions between sub-genomes that have not been completely
396 unzipped during assembly.
397 Overall, the k-mer analysis with Merqury indicates that all the information present in the HiFi
398 reads has been captured in the genome assembly, further suggesting a complete genome.
399

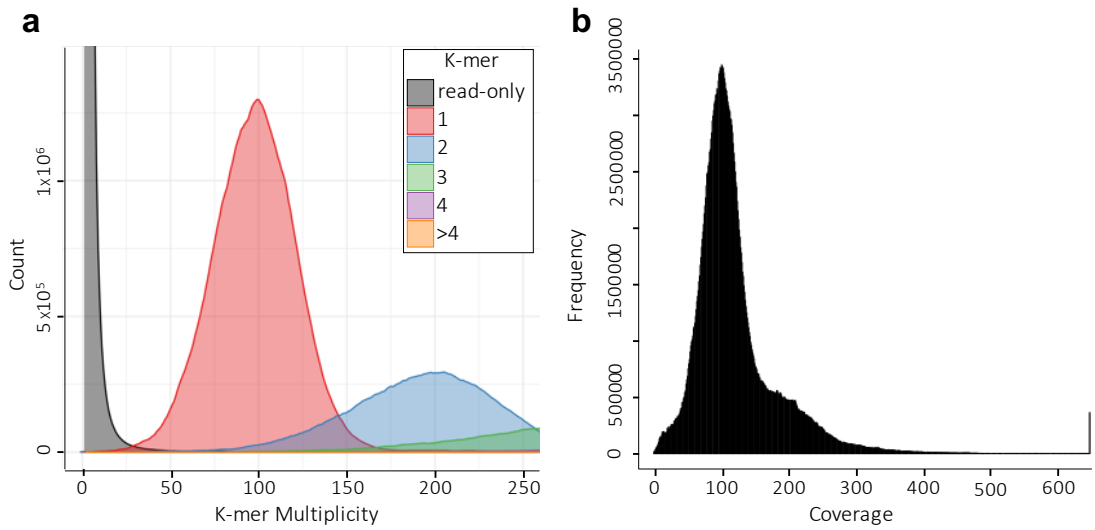


Figure 3. Genome assembly spectra. **A**. The Merqury spectrum plot using DeepConsensus reads tracks the multiplicity of each k-mer detected in the read set. The plot is color-coded according to the number of times a k-mer is found in an assembly. **B**. Bedtools per-base reports coverage for the assembly. The three *M. enterolobii* AAB subgenomes were effectively separated during assembly using Peregrine.

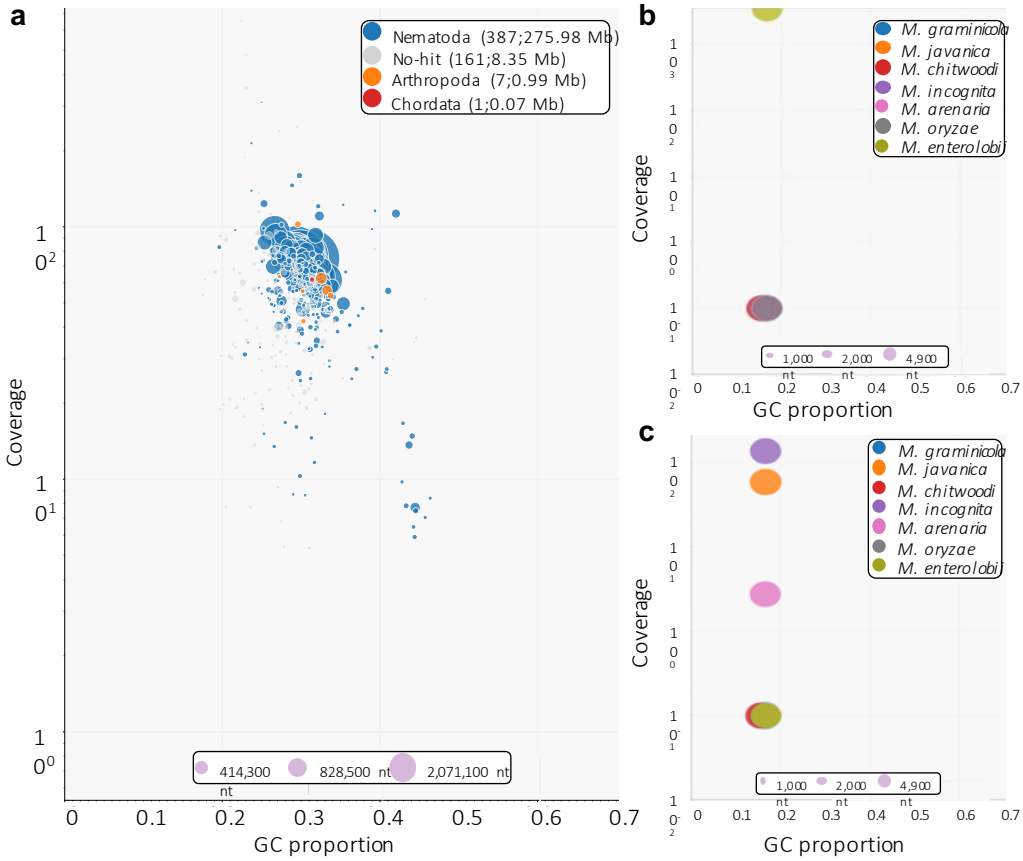
Validating species identity and purity

We confirmed the purity of the *M. enterolobii* population E1834 and the correct species identification by using, first, the Blobtools[26] pipeline. The pipeline generated BlobPlots, which are two-dimensional plots depicting contigs presented as circles, whose diameters are proportional to the sequence length and are colored based on their taxonomic affiliation, determined by the BLAST similarity search results against the NCBI nt database[43]. The relative positions of the circles are according to their GC content and coverage by the long reads. Following the removal of contaminant contigs, the resulting BlobPlot is shown in Figure 4a. Any contigs lacking taxonomic annotation are labeled as 'no-hit'. For the non-Nematoda contigs falling perfectly inside the range of *M. enterolobii* GC content estimates[18] (around 30%), a manual verification was conducted and eight of these contigs were kept. The proposed assignments from Blobtools were disregarded. Instead, for each of them we retained the highest-ranking result proposed by BLAST if the calculated percentage of identity surpassed 90%, the e-value did not exceed $1e^{-50}$, and the taxID belonged to the Nematoda phylum. This resulted in 556 final contigs for this assembly.

The Blobtools pipeline is a valuable tool for detecting possible contaminations in a genome assembly, especially those originating from distant species of different phyla. However, if the contamination comes from a closely related species with a comparable GC content or has been

425 sequenced at a similar coverage, the classical approach will not detect a contamination. For
426 this reason, we made slight adjustments to the methodology (Methods) to achieve a
427 taxonomic classification based on different species within the Nematoda phylum, instead of
428 between phyla only (Fig. 4b, 4c). We focused our analysis on different species within the
429 Meloidogyne genus because (i) they are difficult to differentiate based on the morphology, (ii)
430 they live in the same environment, (iii) they have similar GC content. Therefore, a non-
431 negligible possibility for undetected contamination exists.
432 Using this modified BlobTools methodology, on the *M. enterolobii* population E1834 we have
433 sequenced, we observed that the *M. enterolobii* reference mitochondrial genome from the
434 NCBI was highly covered whereas all the other mitochondrial genomes from the other
435 Meloidogyne species were not covered by our long reads. Hence, no evidence for
436 contamination by other Meloidogyne species was found in the E1834 population (Fig. 4b).
437 For comparison, this method was applied to the previous long-read genome of *M.*
438 *enterolobii*[20], and surprisingly, it was found to be heavily contaminated by another
439 *Meloidogyne* (Fig. 4c). Specifically, the *M. enterolobii* mitochondrial genome was not covered
440 by the previous long reads while those of *M. incognita*, *M. javanica* and *M. arenaria* were all
441 substantially covered. Approximately 60%, 30%, and 10% of the mitochondrial reads aligned
442 with these mitochondrial genomes, respectively. Although this adjusted Blobtools approach
443 suggested contamination of the previous Mma-II Swiss population from other root-knot
444 nematodes, this alone was not sufficient to discriminate between these three closely related
445 species.
446 Consequently, we combined this approach with SCAR markers. All the pairs of primers for the
447 SCAR marker of the four Meloidogyne species of interest were aligned to the previous and
448 current assemblies of *M. enterolobii*. Both for the L30 population of Burkina Faso and the
449 E1834 population from Puerto Rico sequenced here, the pair of primers for the *M. enterolobii*
450 SCAR marker matched the genome assemblies with 100% identity in the correct orientation
451 on one single contig. This allowed identification of a virtual amplified sequence of 537bp,
452 which is consistent with the ~520bp estimated PCR product on the electrophoresis gel in
453 Tigano et al[28]. In contrast, neither the *M. enterolobii* SCAR primers nor the reconstructed
454 corresponding PCR product matched the previous Mma-II genome assembly, confirming the
455 genome was probably not *M. enterolobii*. To further determine the possible source of
456 contamination, we aligned the pairs of primers of the *M. incognita*, *M. javanica* and *M.*
457 *arenaria* SCAR markers on the Mma-II genome. The *M. incognita* pair of primers matched
458 perfectly on this previous assembly in the correct orientation and allowed reconstructing a
459 virtual PCR product of 1192 bp, consistent with the estimated size of the PCR product of ~1,200

460 bp for *M. incognita*[60]. Neither the pair of *M. incognita* primers nor the reconstructed PCR
461 product matched the L30 or E1834 genome assemblies, and none of the *M. javanica* or *M.*
462 *arenaria* pairs of primers matched any of the previously published or current *M. enterolobii*
463 genomes.
464 Therefore, we can conclude that although no trace of contamination by closely related
465 Meloidogyne species could be identified in the L30 or E1834 genome, there is clear evidence
466 that the Mma-II population had been contaminated and replaced by *M. incognita*.
467 The combination of SCAR marker analysis and a modification of Blobtools, specifically for
468 'mitochondrion', has resulted in a powerful tool for the examination and the verification of
469 species purity.
470



471
472 **Figure 4.** BlobPlot of different Meloidogyne genome assemblies. **A**. Blobplot showing
473 taxonomic affiliation at the phylum rank level for the E1834 population of *M. enterolobii*. After
474 removing contamination and mitochondrion, 556 contigs were left. The average GC content
475 for *M. enterolobii* is equal to 30 $\pm$ 0.042. **B**. Mitochondrion only for the E1834 population,
476 after removing contamination. No sign of other Nematoda within the assembly. **C**.
477 Mitochondrion only for the previous *M. enterolobii* reference population provided by

478 Koutsovoulos et al.[20]. No sign of *M. enterolobii* within the assembly. This BlobPlot revealed a
479 contamination by other Meloidogyne spp..
480
481 **Genome completeness assessment**
482 To evaluate the completeness of our genome assembly in terms of expected gene content
483 among related species, we benchmarked nearly universal single-copy orthologs (BUSCO[68]
484 v5.2.2) by using the eukaryote_odb10 lineage dataset in fast mode. Despite the presence of a
485 nematode dataset in BUSCO, it only contains seven species and none of them belong to the
486 same clade as the root-knot nematodes. Therefore, we decided to use the more
487 comprehensive Eukaryotic dataset, which encompasses 70 species. This procedure generates
488 a report that indicates the number of genes that are universally or mostly conserved within
489 the assembly and classifies them into several groups: complete, fragmented, single-copy, or
490 duplicated. The results show that 71.4% (182/255) of BUSCO genes are complete and 12.5%
491 are fragmented. This is a substantial improvement compared to the previously available
492 assemblies (Table 5). Indeed, the Burkina Faso isolate of *M. enterolobii* reached eukaryotic
493 BUSCO completeness score of 59.2% while the Mma-II assembly contaminated by *M. incognita*
494 reached 69.4%.
495 BUSCO is a valuable and robust tool for assessing completeness in a genome assembly in terms
496 of a widely conserved gene set. Nevertheless, in the case of less studied species, the analysis
497 may lack precision if the newly assembled genome comprises variations not included in the
498 initial BUSCO gene set, such as true copy number or sequence variants[40]. We then used
499 Merqury once more to identify any copy-number errors and measure completeness and base
500 accuracy via k-mers. Consequently, Merqury determined the proportion of reliable k-mers in
501 the sequencing sample that were detected in the assembly, resulting in a completeness score
502 of 99.60%. To establish Merqury's base accuracy score, a binomial model for k-mer survival
503 was employed, resulting in a Qscore of 65.70. Higher Qscores indicate a more precise
504 consensus. For instance, Q30 corresponds to an accuracy of 99.9%, Q40 to 99.99%, and so on.
505 In contrast, Burkina Faso and Swiss isolates have Qscores of 55.12 and 29.93, based
506 respectively on their own reads.
507
508
509

BUSCO Categories	*M. enterolobii* (L30[69])	*M. enterolobii** (Swiss[20])	*M. enterolobii* (E1834) **
Complete	59.2% (151)	69.4% (177)	71.4% (182)

Single-copy	29.8% (76)	12.5% (32)	14.9% (38)
Duplicated	29.4% (75)	56.9% (145)	56.5% (144)
Fragmented	18.4% (47)	13.3% (34)	12.5% (32)
Missing	22.4% (57)	17.3% (44)	16.1% (41)

510

511 Table 5. Ortholog BUSCO completeness analysis for different *M. enterolobii* using lineage
512 dataset eukaryota_odb10. *Population contaminated by *M. incognita* *This work.

513

514 **Gene prediction**

515 Using the automated Eugene-EP pipeline, a total of 49,870 genes were predicted, with 45,924
516 being protein-coding genes and 3,946 being non-protein-coding genes such as rRNA, tRNA,
517 and splice leader genes. These genes cover 84 Mb (approximately 29.48%) of the genome
518 assembly length, with the exons spanning 44.51 Mb (around 15.60 %). On average, 5.26 exons
519 are predicted per gene, and the gene length varies from a minimum of 150 bp to a maximum
520 of 35,976 bp. The mean GC content is higher in either the protein-coding region (35.19%) or
521 in the non-protein-coding gene regions (44.19%) compared to that of the whole genome
522 (30.34%).

523

524 **Confirmation of genome structure and ploidy level**

525 Although the *M. enterolobii* population E1834 genome has been predicted as a putative
526 triploid based on k-mer analyses and Smudgeplot, it is important to further confirm the ploidy
527 level of the genome assembly after annotation. The use of MCScanX reveals that a majority of
528 gene duplicates create whole duplicated blocks, rather than dispersed independent
529 duplications. Following the classification established by the duplicate_gene_classifier program
530 implemented in the MCScanX package, 39,532 of the protein-coding genes (around 86.10%)
531 are predicted to be duplicated at least once. As evidenced by Table 6, a majority of these
532 coding genes (75.6%) show a duplication depth of two (meaning for these genes, two other
533 copies exist), further reinforcing the idea that the genome is triploid. Furthermore, it was
534 found that 69.76% of the protein-coding genes fall under the whole-genome duplication
535 category of MCScanX, forming 516 syntenic blocks of collinear genes (see Fig. 5 for
536 visualization of multiple syntenic blocks between different contigs). In addition, 12.61% of the
537 genes are classified as dispersed duplicates, while 2.18% and 1.53% constitute proximal and
538 tandem duplicates, respectively. These findings strongly suggest that the genome of *M.*
539 *enterolobii* is triploid, confirming SmudgePlot results.

540

Duplication depth	0	1	2	3	4	5+
Gene numbers	1992	4917	34720	2700	769	826
Percentage (%)	4.34	10.71	75.60	5.88	1.67	1.80

Table 6. Duplicate gene classifier program of MCScanX for a self-comparison of *M. enterolobii*. Genes with a duplication depth of 0 are not duplicated, while a depth of 1 indicates a maximum of one copy, a depth of 2 indicates two copies, and so forth.

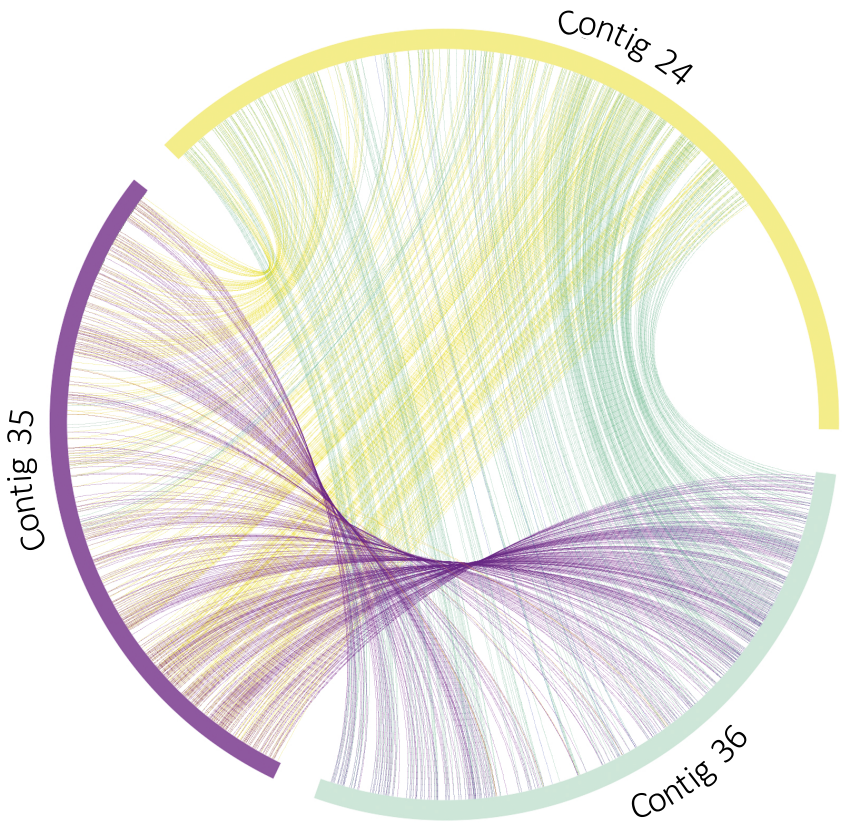


Figure 5. *M. enterolobii* exhibits a triploid genome. The circle plot produced by MCScanX shows collinear gene pairs forming homologous duplicated regions between three contigs. All the collinear gene pairs are linked with different curved colored lines between and within each contig.

Mitochondrial genome assembly and annotation

Using the Aladin package[44], the mitochondrial genome of the *Meloidogyne enterolobii* population E1834 has been assembled and spanned a length of 19,193 bp with a GC content of 17.2% (Fig. 6). We have retrieved and annotated all the mitochondrially encoded subunits involved in the Mitochondrial respiratory chain, including the seven core subunits of the complex I, the cytochrome b of the complex III, the three cytochrome c, and the ATP synthase.

We also obtained a full set of tRNAs, among which were in multiple copies (Ala, Ser, Leu and Asn) as well as ribosomal RNAs (rrnS, rrn12 and rrn16).

When blasted against the NCBI nt database, the reconstructed mitochondrial genome of the *Meloidogyne enterolobii* population E1834 returned as first hit the complete mitochondrial reference genome of *M. enterolobii*[45], with 99.526% identity and an alignment length of 13,067 bp, as the primary hsp. The second-best hit corresponds to an incomplete mitochondrial genome from an *M. enterolobii* isolate discovered on sweet potatoes in the state of Carolina in the USA (GenBank: MW246173.1).

In contrast, reconstruction of the mitochondrial genome using Aladin on the Swiss population Mma-II yielded a ~23kb genome which returned as first hit the *M. incognita* reference mitochondrial genome[58] with >99% identity covering >97% of the query while the *M. enterolobii* reference mitochondrial genome only emerged as the fifth hit with only 87% identity covering 78% of the length.

These results further confirm the E1834 population we have sequenced is indeed *M. enterolobii.*

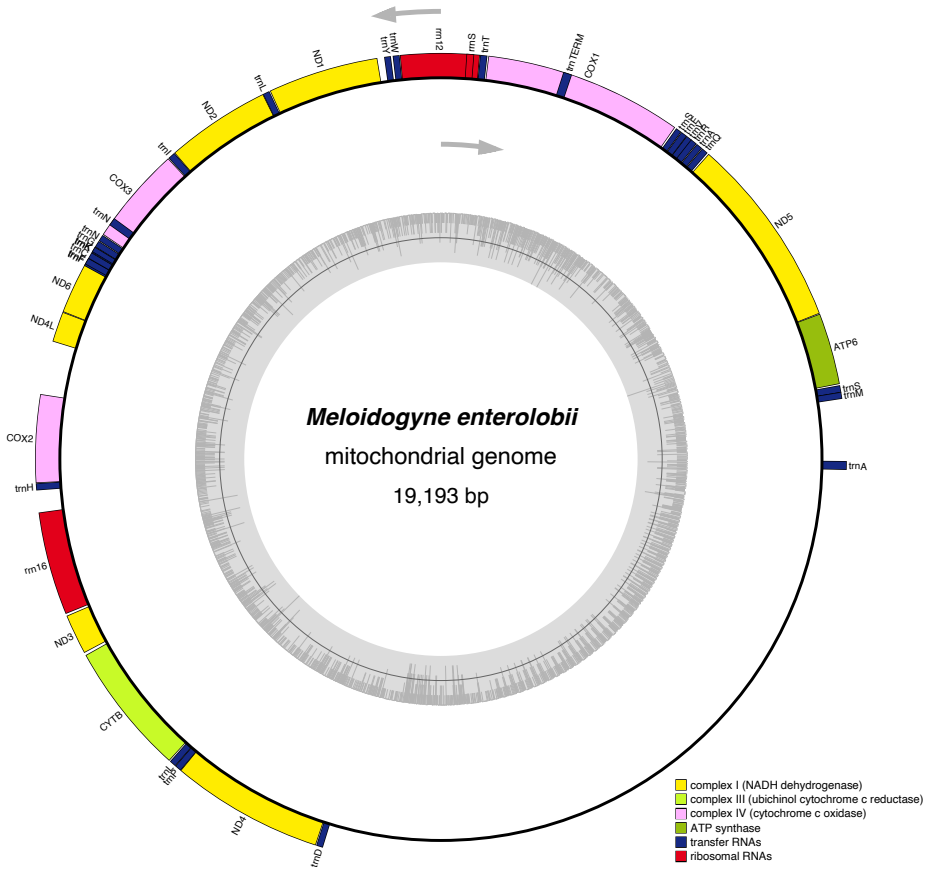


Figure 6. Mitochondrial genome organization of *M. enterolobii*. The inner circle displays the GC content while grey arrows denote the transcription direction. The rRNAs and tRNAs are

577 respectively colored in red and blue. The various complexes of the CRM are represented in
578 yellow, light green, pink, and dark green.
579

580 ## Acknowledgements

581 We are grateful to the colleagues from the Netherlands Institute for Vectors, Invasive Plants
582 and Plant Health (NIVIP) for providing the *M. enterolobii* population E1834 for this study. We
583 thank the genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul,
584 https://doi.org/10.15454/1.5572369328961167E12) for providing computing resources. We
585 are grateful to the bioinformatics and genomics platform, BIG, Sophia Antipolis (ISC plantBIOs,
586 https://doi.org/10.15454/qyey-ar89) for computing and storage resources. We thank Claire
587 Caravel for help in providing the primer sequences for SCAR identification of *Meloidogyne*
588 species. Our *M. enterolobii* genome research is financially supported by a Franco-German
589 bilateral grant ANR-DFG "AEGONE", reference ANR-19-CE35-0017 and reference No
590 431627824.
591

592 ## Author contributions

593 E.G.J.D. and S.K. conceived the research idea and acquired the funding. S.K. supervised all the
594 nematode rearing and DNA extraction experiments, contributed to manuscript writing and
595 reviewing. E.G.J.D. supervised all the bioinformatics analyzes, performed SCAR marker virtual
596 PCRs, contributed to manuscript writing and reviewing. M.P. performed reads processing,
597 genome assembly, contamination and purity check, completeness assessment, ploidy and
598 genome size and structure estimation and wrote the manuscript. H.G. generated single egg
599 mass lines, performed maintenance of the nematode collection, DNA extraction experiments,
600 and contributed to manuscript writing. C.R. performed gene prediction and wrote the
601 corresponding method section. M.S., C.L.R., and J.L. performed library preparation, PacBio HiFi
602 sequencing, and preliminary QC checks and wrote the corresponding method section.
603

604 ## Competing interests

605 The authors declare no competing interests.
606

607 ## References

608 1. Jones, J. T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology.
609 *Molecular Plant Pathology* **14**, 946–961 (2013).
610 2. Santos, D., Abrantes, I. & Maleita, C. The quarantine root-knot nematode Meloidogyne

611 enterolobii – a potential threat to Portugal and Europe. *Plant Pathology* **68**, 1607–1615 (2019).
612 3. Moens, M., Perry, R. N. & Starr, J. L. Meloidogyne species - a diverse group of novel
613 and important plant parasites. in (eds. Perry, R. N., Starr, J. L. & Moens, M.) 1–17 (CABI
614 International, Wallingford, Oxon (CABI), 2009).
615 4. Kiewnick, S., Dessimoz, M. & Franck, L. Effects of the Mi-1 and the N root-knot
616 nematode-resistance gene on infection and reproduction of Meloidogyne enterolobii on
617 tomato and pepper cultivars. *J Nematol* **41**, 134–139 (2009).
618 5. Elling, A. A. Major Emerging Problems with Minor Meloidogyne Species.
619 *Phytopathology®* **103**, 1092–1102 (2013).
620 6. Yang, B. & Eisenback, J. D. Meloidogyne enterolobii n. sp. (Meloidogynidae), a Root-
621 knot Nematode Parasitizing Pacara Earpod Tree in China. *J Nematol* **15**, 381–391 (1983).
622 7. Rammah, A. & Hirschmann, H. Meloidogyne mayaguensis n. sp. (Meloidogynidae), a
623 Root-knot Nematode from Puerto Rico. *J Nematol* **20**, 58–69 (1988).
624 8. Karssen, G., Liao, J., Kan, Z., van Heese, E. Y. & den Nijs, L. J. On the species status of
625 the root-knot nematode Meloidogyne mayaguensis Rammah & Hirschmann, 1988. *Zookeys*
626 67–77 (2012) doi:10.3897/zookeys.181.2787.
627 9. Xu, J., Liu, P., Meng, Q. & Long, H. Characterisation of Meloidogyne species from China
628 using Isozyme Phenotypes and Amplified Mitochondrial DNA Restriction Fragment Length
629 Polymorphism. *European Journal of Plant Pathology* **110**, 309–315 (2004).
630 10. Sikandar, A., Jia, L., Wu, H. & Yang, S. Meloidogyne enterolobii risk to agriculture, its
631 present status and future prospective for management. *Front. Plant Sci.* **13**, 1093657 (2023).
632 11. Castagnone-Sereno, P. Meloidogyne enterolobii (= M. mayaguensis): profile of an
633 emerging, highly pathogenic, root-knot nematode species. *Nematology* **14**, 133–138 (2012).
634 12. Castillo, P. & Castagnone-Sereno, P. Meloidogyne enterolobii (Pacara earpod tree
635 root-knot nematode). *CABI Compendium* **CABI Compendium**, 33238 (2020).
636 13. Schwarz, T., Li, C., Ye, W. & Davis, E. Distribution of Meloidogyne enterolobii in Eastern
637 North Carolina and Comparison of Four Isolates. *Plant Health Progress* **21**, 91–96 (2020).
638 14. Philbrick, A. N., Adhikari, T. B., Louws, F. J. & Gorny, A. M. Meloidogyne enterolobii, a
639 Major Threat to Tomato Production: Current Status and Future Prospects for Its Management.
640 *Frontiers in Plant Science* **11**, (2020).
641 15. THE EUROPEAN COMMISSION. COMMISSION IMPLEMENTING REGULATION (EU)
642 2021/2285 of 14 December 2021 amending Implementing Regulation (EU) 2019/2072 as
643 regards the listing of pests, prohibitions and requirements for the introduction into, and
644 movement within, the Union of plants, plant products and other objects, and repealing
645 Decisions 98/109/EC and 2002/757/EC and Implementing Regulations (EU) 2020/885 and (EU)

646 2020/1292. *Official Journal of the European Union* doi:10.2903/j.efsa.2011.2186.
647 16. Blok, V. C. & Powers, T. O. Biochemical and molecular identification. *Root-knot*
648 *nematodes* 98–118 (2009) doi:10.1079/9781845934927.0098.
649 17. Min, Y. Y., Toyota, K. & Sato, E. A novel nematode diagnostic method using the direct
650 quantification of major plant-parasitic nematodes in soil by real-time PCR. *Nematology* **14**,
651 265–276 (2012).
652 18. Szitenberg, A. *et al.* Comparative Genomics of Apomictic Root-Knot Nematodes:
653 Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biol Evol* **9**, 2844–2861 (2017).
654 19. Blanc-Mathieu, R. *et al.* Hybridization and polyploidy enable genomic plasticity
655 without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics* **13**, e1006777
656 (2017).
657 20. Koutsovoulos, G. D. *et al.* Genome assembly and annotation of Meloidogyne
658 enterolobii, an emerging parthenogenetic root-knot nematode. *Sci Data* **7**, 324 (2020).
659 21. Kiewnick, S., Karssen, G., Brito, J. A., Oggenfuss, M. & Frey, J.-E. First Report of Root-
660 Knot Nematode Meloidogyne enterolobii on Tomato and Cucumber in Switzerland. *Plant*
661 *Disease* **92**, 1370–1370 (2008).
662 22. Marx, V. Method of the year: long-read sequencing. *Nat Methods* **20**, 6–11 (2023).
663 23. Mota, A. P. Z. *et al.* Unzipped genome assemblies of polyploid root-knot nematodes
664 reveal unusual and clade-specific telomeric repeats. *Nat Commun* **15**, 773 (2024).
665 24. Dai, D. *et al.* Unzipped chromosome-level genomes reveal allopolyploid nematode
666 origin pattern as unreduced gamete hybridization. *Nat Commun* **14**, 7156 (2023).
667 25. Gerič Stare, B., Strajnar, P., Susič, N., Urek, G. & Širca, S. Reported populations of
668 Meloidogyne ethiopica in Europe identified as Meloidogyne luci. *Plant Disease* **101**, 1627–
669 1632 (2017).
670 26. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. Preprint
671 at https://doi.org/10.12688/f1000research.12232.1 (2017).
672 27. Kiewnick, S., Frey, J. E. & Braun-Kiewnick, A. Development and Validation of LNA-
673 Based Quantitative Real-Time PCR Assays for Detection and Identification of the Root-Knot
674 Nematode Meloidogyne enterolobii in Complex DNA Backgrounds. *Phytopathology®* **105**,
675 1245–1249 (2015).
676 28. Tigano, M. *et al.* Genetic diversity of the root-knot nematode Meloidogyne enterolobii
677 and development of a SCAR marker for this guava-damaging species. *Plant Pathology* **59**,
678 1054–1061 (2010).
679 29. Ye, W., Zeng, Y. & Kerns, J. Molecular Characterisation and Diagnosis of Root-Knot
680 Nematodes (Meloidogyne spp.) from Turfgrasses in North Carolina, USA. *PLOS ONE* **10**,

681 e0143556 (2015).
682 30. Gómez-González, G. *et al.* Meloidogyne enterolobii egg extraction in NaOCl versus
683 infectivity of inoculum on cucumber. *J Nematol* **53**, e2021-57 (2021).
684 31. Jenkins, W. R. A rapid centrifugal-flotation technique for separating nematodes from
685 soil. Plant Dis. Rep. 48:692, 1964.
686 32. Baid, G. *et al.* DeepConsensus improves the accuracy of sequences with a gap-aware
687 sequence transformer. *Nat Biotechnol* **41**, 232–238 (2023).
688 33. Töpfter, A. PacificBiosciences/ACTC: Align subreads to CCS reads. Available at:
689 https://github.com/PacificBiosciences/actc. PacBio (2022).
690 34. Kokot, M., Długosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer
691 statistics. *Bioinformatics* **33**, 2759–2761 (2017).
692 35. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and
693 Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).
694 36. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of
695 occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
696 37. Sim, S. B., Corpuz, R. L., Simmonds, T. J. & Geib, S. M. HiFiAdapterFilt, a memory
697 efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi
698 reads and their negative impacts on genome assembly. *BMC Genomics* **23**, 157 (2022).
699 38. Chin, J. Peregrine-2021: A faster and minimum genome assembler. (2023).
700 39. Chin, C.-S. & Khalak, A. Human Genome Assembly in 100 Minutes. 705616 Preprint at
701 https://doi.org/10.1101/705616 (2019).
702 40. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality,
703 completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245
704 (2020).
705 41. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**,
706 4572–4574 (2021).
707 42. McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of
708 sequence analysis tools. *Nucleic Acids Res* **32**, W20–W25 (2004).
709 43. Sayers, E. W. *et al.* Database resources of the National Center for Biotechnology
710 Information. *Nucleic Acids Res* **50**, D20–D26 (2021).
711 44. Koutsovoulos, G. GDKO/Aladin: Aladin (mitochondrial circular DNA reconstitution).
712 Available at: https://github.com/GDKO/aladin. GitHub (2021).
713 45. Humphreys-Pereira, D. A. & Elling, A. A. Meloidogyne enterolobii mitochondrion,
714 complete genome. Available at: https://www.ncbi.nlm.nih.gov/nuccore/NC_026555.1. NCBI.
715 National Center for Biotechnology Information (2015).

716 46. Tillich, M. *et al.* GeSeq – versatile and accurate annotation of organelle genomes.
717 *Nucleic Acids Research* **45**, W6–W11 (2017).
718 47. Chan, P. P. & Lowe, T. M. tRNAscan-SE: Searching for tRNA genes in genomic
719 sequences. *Methods Mol Biol* **1962**, 1–14 (2019).
720 48. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes
721 in nucleotide sequences. *Nucleic Acids Research* **32**, 11–16 (2004).
722 49. Laslett, D. & Canbäck, B. ARWEN: a program to detect tRNA genes in metazoan
723 mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175 (2008).
724 50. Sallet, E., Gouzy, J. & Schiex, T. EuGene: An Automated Integrative Gene Finder for
725 Eukaryotes and Prokaryotes. in *Gene Prediction: Methods and Protocols* (ed. Kollmar, M.) 97–
726 120 (Springer, New York, NY, 2019). doi:10.1007/978-1-4939-9173-0_6.
727 51. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite − a
728 comprehensive resource for helminth genomics. *Mol Biochem Parasitol* **215**, 2–10 (2017).
729 52. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic Acids*
730 *Res* **46**, 2699 (2018).
731 53. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements
732 in eukaryotic genomes. *Mobile DNA* **6**, 11 (2015).
733 54. Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene
734 synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
735 55. Sun, L., Zhuo, K., Lin, B., Wang, H., & Liao, J. Meloidogyne graminicola mitochondrion,
736 complete genome. Available at: https://www.ncbi.nlm.nih.gov/nuccore/NC_056772.1. NCBI.
737 National Center for Biotechnology Information (2014).
738 56. Humphreys-Pereira, D. A. & Elling, A. A. Meloidogyne arenaria mitochondrion,
739 complete genome. Available at: https://www.ncbi.nlm.nih.gov/nuccore/NC_026554.1.
740 National Center for Biotechnology Information (2015).
741 57. Humphreys-Pereira, D. A. & Elling, A. A. Meloidogyne javanica mitochondrion,
742 complete genome. Available at: https://www.ncbi.nlm.nih.gov/nuccore/NC_026556.1.
743 National Center for Biotechnology Information (2015).
744 58. Humphreys-Pereira, D. A. & Elling, A. A. Mitochondrial genomes of Meloidogyne
745 chitwoodi and M. incognita (Nematoda: Tylenchina): Comparative analysis, gene order and
746 phylogenetic relationships with other nematodes. *Molecular and Biochemical Parasitology*
747 **194**, 20–32 (2014).
748 59. Besnard, G. *et al.* On the close relatedness of two rice-parasitic root-knot nematode
749 species and the recent expansion of Meloidogyne graminicola in Southeast Asia. Genes, 10(2),
750 175. https://doi.org/10.3390/genes10020175 (2019).

751 60. Zijlstra, C., Donkers-Venne, D. T. H. M. & Fargette, M. Identification of Meloidogyne
752 incognita, M. javanica and M. arenaria using sequence characterised amplified region (SCAR)
753 based PCR assays. *Nematology* **2**, 847–853 (2000).
754 61. European Nucleotide Archive. https://identifiers.org/ena.embl:PRJEB69523. (2024).
755 62. Poullet Marine & Danchin Etienne G. J. Meloidogyne enterolobii assemblies. DRAFT
756 VERSION. Recherche Data Gouv https://doi.org/10.57745/5MXZSJ (2023).
757 63. Poullet Marine, Danchin Etienne G. J, & Rancurel Corinne. Meloidogyne enterolobii
758 E1834 gene prediction. DRAFT VERSION. Recherche Data Gouv
759 https://doi.org/10.57745/Y0O2LP (2023).
760 64. Poullet Marine & Danchin Etienne G. J. Meloidogyne enterolobii E1834 structural
761 annotation. DRAFT VERSION. Recherche Data Gouv https://doi.org/10.57745/VEKHQS (2023).
762 65. Jaron, K. S. *et al.* Genomic Features of Parthenogenetic Animals. *Journal of Heredity*
763 **112**, 19–33 (2021).
764 66. Liu, B. *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de
765 novo genome projects. Preprint at https://doi.org/10.48550/arXiv.1308.2012 (2020).
766 67. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic
767 features. *Bioinformatics* **26**, 841–842 (2010).
768 68. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update:
769 Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for
770 Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**,
771 4647–4654 (2021).
772 69. Lunt, D. H. Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent
773 hybrid origins. *BMC Evol Biol* **8**, 194 (2008).
774