

Genome sequence of the coffee root-knot nematode *Meloidogyne exigua*

Ngan Thi Phan¹,
Guillaume Besnard²,
Rania Ouazahrour³,
William Solano Sánchez⁴,
Lisa Gil⁵, Sophie Manzi²
and Stéphane Bellaïflore^{1*}

¹PHIM Plant Health Institute,
University of Montpellier, IRD,
CIRAD, INRAE, Institut Agro,
Montpellier, France.

²CNRS-UPS-IRD, UMR5174, EDB,
118 route de Narbonne, Université
Paul Sabatier, 31062 Toulouse,
France.

³University of Rennes 1, UFR SVE,
35065, Rennes, France.

⁴CATIE – Centro Agronómico
Tropical de Investigación y
Enseñanza, Turrialba, Costa Rica.

⁵US 1426, GeT-PlaGe, Genotoul,
INRAE, Castanet-Tolosan, France.

*E-mail: stephane.bellaïflore@ird.fr

This paper was edited by
Dee Denver.

Received for publication
March 17, 2021.

Abstract

Root-knot nematodes (*Meloidogyne* spp.) cause serious damages on most crops. Here, we report a high-quality genome sequence of *Meloidogyne exigua* (population Mex1, Costa Rica), a major pathogen of coffee. Its mitogenome (20,974 bp) was first assembled and annotated. The nuclear genome was then constructed consisting of 206 contigs, with an N50 length of 1.89 Mb and a total assembly length of 42.1 Mb.

Keywords

Genomics, Illumina, Mitogenome, Nanopore sequencing, Nuclear genome, Root-knot nematode.

Root-knot nematodes (RKN) parasitize a wide range of host plants and have a global distribution. They are considered the most important group of plant-parasitic nematodes (Jones et al., 2013). Several *Meloidogyne* species can attack coffee plants, but only *Meloidogyne exigua* (Goeldi, 1892) has a significant impact on coffee production. This pathogen is the most widely distributed nematode in the coffee production areas in Central and South America (Campos and Villain, 2005), with estimated yield losses of up to 45% in the Rio de Janeiro State (Barbosa et al., 2004) and between 15 and 20% in Central America as a whole (Anzueto et al., 1995). Despite these serious impacts on coffee production,

diversity and adaptation of *M. exigua* has been poorly documented, and so far, the only published study on the species was based on isozyme profiles and random amplified polymorphic DNA (RAPD) markers (Muniz et al., 2008). With the advent of high throughput sequencing methods, the analysis of its genome has become possible and may open new avenues for studying its evolutionary history.

Comparative genomics of RKN species has revealed a striking diversity in genome structure (e.g. chromosome counts, ploidy level, duplicated regions, heterozygosity) that might be linked to their different reproductive modes and species origin (Blanc-Mathieu et al., 2017; Castagnone-Sereno

et al., 2013; Jaron et al., 2020; Triantaphyllou, 1985). Interestingly, despite prominent asexual reproduction in several RKN species, various mechanisms can generate genomic variability and may play a major role in their adaptability against different environments and hosts. These include, in particular, horizontal gene transfers (Danchin et al., 2016; Opperman et al., 2008; Phan et al., 2020), insertion of transposable elements (Kozłowski et al., 2020), and gene duplications/deletions (i.e. gene copy number variants; Castagnone-Sereno et al., 2019). *M. exigua* is a successful pathogen on coffee with a parthenogenetic reproduction mode (Triantaphyllou, 1985), and as demonstrated in other RKNs, its

adaptation to various conditions may be also favored by above mentioned mechanisms. Here, we report a high-quality genome assembly of the genome of *M. exigua* population 'Mex1'. The assembly represents a valuable molecular resource for future studies of phylogenomics on *Meloidogyne* species. In particular, this will foster comparative genomics to investigate and understand the evolutionary history of this nematode, the results of which may help in the development of new strategies for its management.

We used long-read Oxford Nanopore Technology (ONT) and short-read Illumina HiSeq sequencing data to generate the genome assembly. The population 'Mex1' was isolated from coffee roots collected in

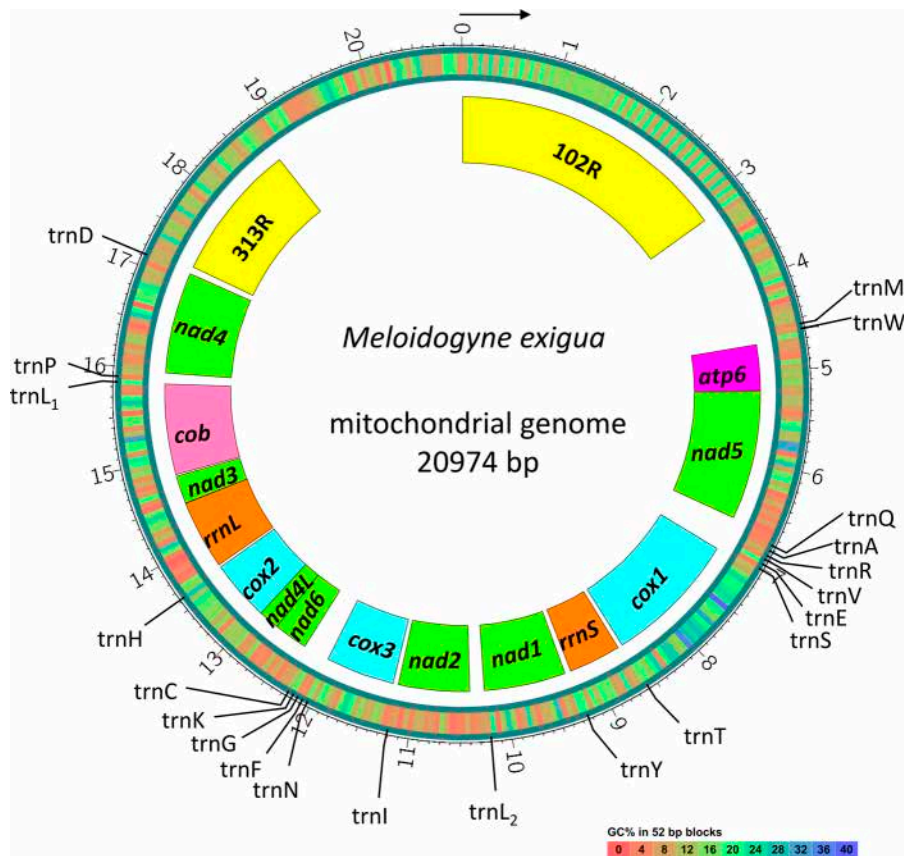


Figure 1: Circular gene map of the complete mitochondrial genome of *M. exigua*. Protein-coding genes, rRNA genes, and repeated regions are represented as boxes. Position of tRNAs are indicated by black lines. The direction of transcription of all genes is the same and indicated by the arrow. Abbreviations of protein coding and rRNA genes are: *nad*i = subunit i of NADH dehydrogenase; *cox*i = subunit i of cytochrome c oxidase; *cob* = cytochrome b; *atp*6 = subunit 6 of ATP-synthase; *rrn*S = small subunit ribosomal RNA (12S); *rrn*L = large subunit ribosomal RNA (16S). tRNA genes (trnX) are named with a single-letter amino acid abbreviation (X) except for those coding for leucine, which are named as L₁ (anticodon uag) and L₂ (uua). Two minisatellite regions, namely 102R and 313R, are composed of 102-bp and 313-bp repeats.

Hacienda Aquiares located in Turrialba, Cartago, Costa Rica (9°56'18.09"N, 83°43'43.86"W). A single juvenile was inoculated and multiplied on tomato (*Solanum lycopersicum* var. Moneymaker). The procedures for sequencing of *M. exigua* genome including nematode extraction and purification, genomic DNA extraction and purification, library preparation, and sequencing processes for the ONT and Illumina platforms were as described by Phan et al. (2020). For ONT sequencing, six micrograms of purified DNA were used to produce 3,150,177 raw reads with a total length of 15.16Gb (N50 length=13.9kb; ca. 150-fold genome coverage). The ONT reads were trimmed to remove adapters using Porechop v.0.2.3 (Wick, 2019). Then, sequence was filtered for quality (Q-score \geq 9) and length ($L\geq$ 500bp) using NanoFilt v.1.1.0 (De Coster et al., 2018). Finally, 13.75Gb of trimmed long reads (coverage of 137x) were selected for further analysis. Reads from the Illumina Technology were obtained from 3 μ g of gDNA using the HiSeq3000 platform as described by Phan et al. (2020). Paired-end reads of 150bp were generated (mean insert size=452bp), yielding 43.08 million reads (64.6 Gb; ca. 153-fold genome coverage). The quality of Illumina raw reads were assessed using FastQC (Andrews, 2010). Spades v.3.14.1 (Bankevich et al., 2012) and Blobtools v.2.1 (Kumar et al., 2013) were used to identify possible contamination; however, no potential contamination was detected. The Skewerv.0.2.2 software (Jiang et al., 2014) was used to trim reads using a minimum quality score of 30 and a minimum read length of 51bp. Finally, the reads were error-corrected using Musket v.1.1 (Liu et al., 2013). Finally, 43.01 million trimmed pair-end reads (64.4Mb, coverage of 152x) were used for the genome assembly.

The mitochondrial genome (mitogenome) of *M. exigua* was de novo assembled using short reads following the experimental procedure described by Besnard et al. (2014). Long reads were used to resolve the repeated sequences. A mitogenome sequence of 20,974bp was constructed with an average coverage of 18,698x. Protein-coding genes and transfer RNAs (tRNAs) were annotated using the prediction pipeline of Mitos (Bernt et al., 2013; Donath et al., 2019) with the invertebrate mitochondrial code. Blastn search against the mitogenomes of *M. graminicola* (NC_024275.1) and *M. chitwoodii* (KJ476150) was used to confirm the prediction and to manually check the position of start/stop codons. Fourteen protein coding genes (*atp6*, *nad5*, *cox1*, *nad1*, *nad2*, *cox3*, *nad6*, *nad4L*, *cox2*, *rrnL*, *nad3*, *cob*, and *nad4*), two ribosomal RNA (rRNA) genes (*rrnS* and *rrnL*), two repeated regions (102R and 313R), and 21 transfer

ribonucleic acid (tRNA) genes (*trnM*, *trnW*, *trnQ*, *trnA*, *trnR*, *trnV*, *trnE*, *trnS*, *trnT*, *trnY*, *trnL*₂, *trnI*, *trnN*, *trnF*, *trnG*, *trnK*, *trnC*, *trnH*, *trnL*₄, *trnP*, and *trnD*) were finally annotated from the mitogenome sequence. The mitogenome structure was visualized using the CIRCOS software (<http://circos.ca/>) (Fig. 1).

The reads that mapped to the mitogenome (with 100% identity; CIGAR=100M) were removed from the cleaned long and short reads datasets and the remaining sequences used for assembly of nuclear genome. The Canuv.1.8 software (Koren et al., 2017) was first used for the assembly. Subsequently, Racon v.1.4.3 (Vaser et al., 2017) and Pilon v.1.23 (Walker et al., 2014) were used to correct bases and homopolymer lengths. Contigs that had low read-coverage (<10x) were eliminated from the assembly to avoid artifacts and possible contamination. Finally, the assembled genome consisted of 206 contigs with a maximum contig length of 3,958Kb and N50 of 1,882Kb (Table 1). The total length of the assembly is 42.10Mb, which matches the estimated haploid genome length of 43.2Mb based on *k*-mer analyses (at *k*=21) using Jellyfish v.1.0 (Marçais and

Table 1. Statistics of the genome assembly for *M. exigua* obtained in our study (with Canu; Koren et al., 2017).

Assembly features	<i>M. exigua</i> genome
Total #scaffolds	260
Total length (bp)	42,101,073
Largest contig (bp)	3,958,915
N50 (bp)	1,882,513
N90 (bp)	1,045,864
L50 (# scaffolds)	10
L90 (#scaffolds)	18
GC (%)	25.55
Mismatches	0
Gaps	0
CEGMA completeness ^a	C:95.97% (C+P: 97.18%)
BUSCO completeness ^b	C:89.4% [S: 89.1%, D: 0.3%]

Notes: ^aC: Complete; C+P: Complete+Partial; ^bC: Complete; S: Complete and single-copy; D: Complete and duplicated.

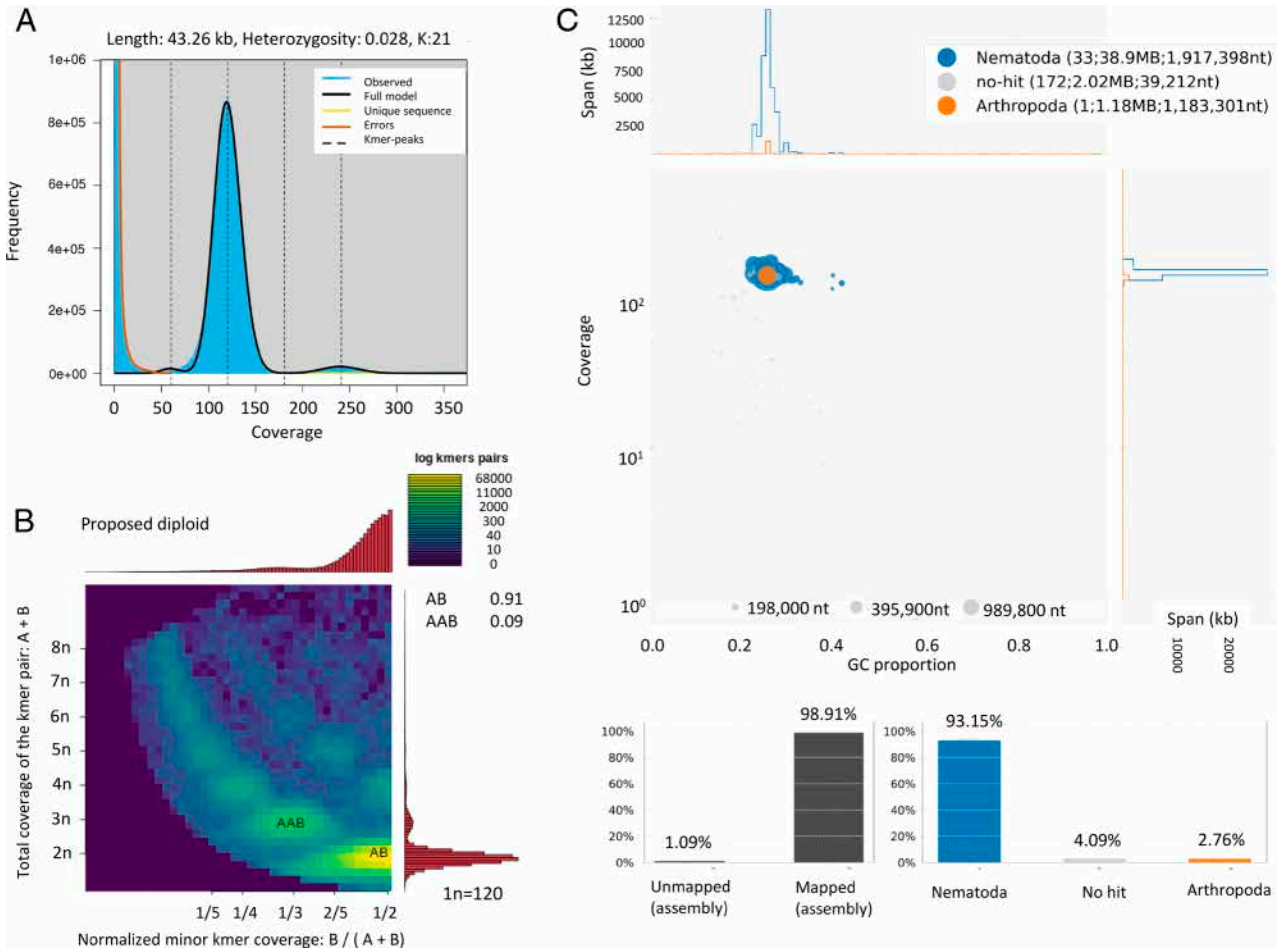


Figure 2: Haploid genome length, genome ploidy estimation and contaminant analysis of the *M. exigua* genome assembly. (A) GenomeScope profile showing estimated genome length of 43.26 Mb and heterozygosity of 0.028% at k -mer=21. (B) Smudge plots showing the coverage and distribution of k -mer pairs that fit to diploid genome model. (C) Blobplot showing the lack of contamination in the final assembly by foreign (non-Nematoda) genetic material.

Kingsford, 2011) and GenomeScope v.2.0 (Vurture et al., 2017) (Table 1 and Fig. 2A). Smudgeplot v.0.1.3 (Ranallo-Benavidez et al., 2020) and KMCv.3.0.0 softwares (Kokot et al., 2017) were used to estimate genome ploidy based on the k -mers counting ($k=21$) of the short-read data. The genome is estimated to be diploid (AB) with heterozygosity of 0.03% (Fig. 2B). Blobtools (Laetsch and Blaxter, 2017) was used to assess contaminant DNA presence on the final genome assembly (Fig. 2C). Most of the genome assembly belong to Nematoda phylum (93.1%; Fig. 2C). One scaffold (1.18Mb) was, however, assigned to the Arthropoda phylum (Fig. 2C). However, sequencing coverage and GC content of this scaffold were similar to other contigs of the genome assembly, and should thus be part of the nematode genome (Fig. 2C). The GC content of the assembled genome was

25.5% (Table 1). The Core Eukaryotic Genes Mapping Approach (CEGMA v.2.5) analysis (Parra et al., 2007) revealed that genome assembly contain 95.75% among 248 Eukaryotic Orthologs. The average number of orthologs per core gene at 1.09 indicated a haploid genome assembly. Besides, the genome assembly was 89.4% complete based on the eukaryote set ($n=303$) of Benchmarking Universal Single-Copy Orthologs (BUSCO v.3.0.2) (Simão et al., 2015). Among available *Meloidogyne* genomes, this new assembly yields the second highest BUSCO completeness (after *Meloidogyne javanica*, summarized in Koutsovoulos et al., 2020) and the second largest N50 length (after *Meloidogyne chitwoodi*; Bali et al., 2021). This reference will assist a range of genetic, genomic, and phylogenetic studies to uncover diversity and evolution of *M. exigua* and other related RKNs.

Data availability and accession number(s)

Procedural information concerning the genome assembly and analysis presented in this paper can be found at the GitHub repository at https://github.com/PhanNgan/genome_assembly_mex. The mitogenome and nuclear genome sequences have been deposited in DDBJ/ENA/GenBank under the accession numbers MZ359281 and JAGUQR000000000, respectively. The nuclear genome version described in this paper is version JAGUQR010000000.

Acknowledgments

This research was funded by the Consultative Group for International Agricultural Research Program on rice-agrifood systems (CRP-RICE, 2017–2022). The sequencing was performed at the GeT-PlaGe core facility, INRAE Toulouse, France. Ngan Thi Phan was supported by a PhD fellowship from French Embassy in Vietnam and by the LABEX Agro coordinated by Agropolis Fondation (ANR-10-LABX-001-01). Guillaume Besnard is member of the EDB lab that is supported by the LABEX TULIP (ANR-10-LABX-0041) and CEBA (ANR-10-LABX-25-01). The authors are grateful to J. Aribi (PHIM) for his technical assistance.

References

- Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data Babraham Bioinformatics, The Babraham Institute, Cambridge, available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed March 15, 2020).
- Anzueto, F., Bertrand, B. and Dufour, M. 1995. “Nemaya”, desarrollo de una variedad porta-injerto resistente a los principales nemátodos de América Central. Programa Cooperativo para el Desarrollo Tecnológico y la Modernización de la Caficultura (PROMECAFE) Boletín PROCAFE-IICA-CATIE, No. 66-67, available at: <http://repositorio.iica.int/handle/11324/7910>.
- Bali, S., Hu, S., Vining, K., Brown, C. R., Majtahedi, H., Zhang, L., Gleason, C. and Sathuvalli, V. 2021. Nematode genome announcement: draft genome of *Meloidogyne chitwoodi*, an economically important pest of potato in the Pacific Northwest IS-MPMI Congress.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* 19:455–77.
- Barbosa, D. H. S. G., Duarte, H., Souza, R., Viana, A. P. and Silva, C. P. 2004. Field estimates of coffee yield losses and damage threshold by *Meloidogyne exigua*. *Nematologia Brasileira* 28:49–54.
- Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsche, G., Pütz, J., Middendorf, M. and Stadler, P. F. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution* 69:313–9.
- Besnard, G., Jühling, F., Chapuis, É., Zedane, L., Lhuillier, É., Mateille, T. and Bellaïre, S. 2014. Fast assembly of the mitochondrial genome of a plant parasitic nematode *Meloidogyne graminicola* using next generation sequencing. *Comptes Rendus Biologies* 337:295–301.
- Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J. -M., Rocha, M. D., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J. -F., Kozłowski, D. K., Cazareth, J., Couloux, A., Silva, C. D., Guy, J., Kim-Jo, Y. -J., Rancurel, C., Schiex, T., Abad, P., Wincker, P. and Danchin, E. G. J. 2017. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLOS Genetics* 13:e1006777.
- Campos, V. P. and Villain, L. 2005. Nematode parasites of coffee and cocoa. In Luc, M., Sikora, R. A. and Bridge, J. (Eds), *Plant parasitic nematodes in subtropical and tropical agriculture*, Wallingford: CAB, pp. 529–79.
- Castagnone-Sereno, P., Mulet, K., Danchin, E. G. J., Koutsovoulos, G. D., Karaulic, M., Rocha, M. D., Bailly-Bechet, M., Pratz, L., Perfus-Barbeoch, L. and Abad, P. 2019. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*. *Molecular Ecology* 28:2559–72.
- Danchin, G. J. E., Guzeeva, A. E., Mantelin, S., Berepiki, A. and Jones, T. J. 2016. Horizontal gene transfer from bacteria has enabled the plant-parasitic nematode *Globodera pallida* to feed on host-derived sucrose. *Molecular Biology and Evolution* 33:1571–9.
- De Coster, W., D’Hert, S., Schultz, D. T., Cruts, M. and Van Broeckhoven, C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–9.
- Donath, A., Jühling, F., Al-Arab, M., Bernhart, S. H., Reinhardt, F., Stadler, P. F., Middendorf, M. and Bernt, M. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research* 47:10543–52.
- Goeldi, E. A. 1892. Relatoria sobre a molestia do cafeeiro na provincia do Rio de Janeiro. *Archivos do Museu Nacional do Rio de Janeiro* 8:1–121.
- Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M. and Schwander, T. 2020. Genomic features of parthenogenetic animals. *Journal of Heredity* 1:19–33.

- Jiang, H., Lei, R., Ding, S.-W. and Zhu, S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:182.
- Jones, J. T., Haegeman, A., Danchin, E. G. J., Gaur, H. S., Helder, J., Jones, M. G. K., Kikuchi, T., Manzanilla-López, R., Palomares-Rius, J. E., Wesemael, W. M. L. and Perry, R. N. 2013. Top 10 plant-parasitic nematodes in molecular plant pathology. *Molecular Plant Pathology* 14:946–61.
- Kokot, M., Długosz, M. and Deorowicz, S. 2017. KMC 3: counting and manipulating *k*-mer statistics. *Bioinformatics* 33:2759–61.
- Koutsovoulos, G. D., Pouillet, M., Ashry, A. E., Kozłowski, D. K., Sallet, E., Rocha, M. D., Perfus-Barbeoch, L., Martin-Jimenez, C., Frey, J. E., Ahrens, C. H., Kiewnick, S. and Danchin, E. G. J. 2020. The polyploid genome of the mitotic parthenogenetic root-knot nematode *Meloidogyne enterolobii*. *Scientific Data* 7:324.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H. and Phillippy, A. M. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research* 27:722–36.
- Kozłowski, D. K., Hassanaly-Goulamhousen, R., Rocha, M. D., Koutsovoulos, G. D., Bailly-Bechet, M. and Danchin, E. G. J. 2020. Transposable elements are an evolutionary force shaping genomic plasticity in the parthenogenetic root-knot nematode *Meloidogyne incognita*. *bioRxiv*2020.04.30.069948.
- Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M. and Blaxter, M. 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Frontiers in Genetics* 4:237.
- Laetsch, D. R. and Blaxter, M. L. 2017. BlobTools: interrogation of genome assemblies. *F1000Research* 6:1287.
- Liu, Y., Schröder, J. and Schmidt, B. 2013. Musket: a multistage *k*-mer spectrum-based error corrector for Illumina sequence data. *Bioinformatics* 29:308–15.
- Marçais, G. and Kingsford, C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* 27:764–70.
- Muniz de, M. F., Carneiro, R., Almeida, M. R., Campos, V. P., Castagnone-Sereno, P. and Castro, J. M. C. 2008. Diversity of *Meloidogyne exigua* (Tylenchida: Meloidogynidae) populations from coffee and rubber tree. *Nematology* 10:897–910.
- Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T. D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B. R., Thomas, V. P. and Windham, E. 2008. Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences of the United States of America* 105:14802–7.
- Parra, G., Bradnam, K. and Korf, I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* 23:1061–7.
- Phan, N. T., Orjuela, J., Danchin, E. G. J., Klopp, C., Perfus-Barbeoch, L., Kozłowski, D. K., Koutsovoulos, G. D., Lopez-Roques, C., Bouchez, O., Zahm, M., Besnard, G. and Bellafiore, S. 2020. Genome structure and content of the rice root-knot nematode (*Meloidogyne graminicola*). *Ecology and Evolution* 10:11006–21.
- Ranallo-Benavidez, T. R., Jaron, K. S. and Schatz, M. C. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11:1432.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–2.
- Triantaphyllou, A. C., Sasser, J. N. and Carter, C. C. 1985. Cytogenetics, cytotaxonomy and phylogeny of root-knot nematodes. An advanced treatise on *Meloidogyne* biology and control Raleigh: North Carolina State University, 113–26.
- Vaser, R., Sovic, I., Nagarajan, N. and Sikic, M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research* 27:737–46.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J. and Schatz, M. C. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* 33:2202–4.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S.K. and Earl, A. M. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963.
- Wick, R. 2019. Porechop [C++], available at: <https://github.com/rrwick/Porechop> (original work published 2017).