

1 Article

2 **First Draft Genome of the Trypanosomatid** 3 *Herpetomonas muscarum ingenoplastis* through 4 **MinION Oxford Nanopore Technology and Illumina** 5 **Sequencing**

6 **Claudia M. d'Avila-Levy^{1,2*}, Bertrand Bearzatto³, Jérôme Ambroise³, Raphaël Helaers², Anzhelika**
7 **Butenko^{4,5}, Vyacheslav Yurchenko^{5,6}, Karina A. Morelli^{1,7}, Helena L. C. Santos¹, Pascal**
8 **Brouillard², Philippe Grellier⁸, Jean-Luc Gala³, Miikka Vikkula²**

9 ¹ Coleção de Protozoários, Laboratório de Estudos Integrados em Protozoologia, Instituto Oswaldo Cruz,
10 Fundação Oswaldo Cruz, 21040-360 Rio de Janeiro, Brazil; davila.levy@ioc.fiocruz.br

11 ² de Duve Institute, University of Louvain, B-1200, Brussels, Belgium; miikka.vikkula@uclouvain.be

12 ³ Centre de Technologies Moléculaires Appliquées, University of Louvain, B-1200, Brussels, Belgium; jean-
13 luc.gala@uclouvain.be

14 ⁴ Institute of Parasitology, Biology Centre, Czech Academy of Sciences, 37005 České Budějovice (Budweis),
15 Czech Republic; rolando24@yandex.ru

16 ⁵ Life Science Research Centre, Faculty of Science, University of Ostrava, 71000 Ostrava, Czech Republic;
17 vyacheslav.yurchenko@osu.cz

18 ⁶ Martsinovskiy Institute of Medical Parasitology, Sechenov University, Moscow, 119435 Russia;

19 ⁷ Instituto de Biologia Roberto Alcântara Gomes, Departamento de Ecologia, Universidade Estadual do Rio
20 de Janeiro, 20550-900, Rio de Janeiro, Brasil; kmorelli@ioc.fiocruz.br

21 ⁸ Unité Molécules de Communication et Adaptation des Microorganismes (UMR 7245 CNRS MCAM),
22 Muséum National d'Histoire Naturelle, Sorbonne Universités, 75005 Paris, France ; grellier@mnhn.fr

23
24 * Correspondence: davila.levy@ioc.fiocruz.br (D.L.); Tel.: +55-21-2562-1014 (CMdL)

25

26 **Abstract:** We presented here the first draft genome sequence of the trypanosomatid *Herpetomonas*
27 *muscarum ingenoplastis*. This parasite was isolated repeatedly in the black blowfly, *Phormia regina*.
28 This is the first draft genome of a flagellate from the phylogenetically distinct clade of
29 Trypanosomatidae.

30 **Keywords:** genome assembly; monoxenous trypanosomatids; insect trypanosomatids;
31 Trypanosomatidae; whole genome

32

33 1. Introduction

34 The family Trypanosomatidae (Kinetoplastea: Trypanosomatida) comprises parasites of
35 vertebrates, invertebrates or plants [1]. Chagas disease, leishmaniasis and human African
36 trypanosomiasis are human diseases caused by *Trypanosoma cruzi*, *Leishmania* spp. and *Trypanosoma*
37 *brucei* sensu lato, respectively [2]. These parasites affect about 22 million people worldwide and
38 alternate their life cycle between an insect vector and a mammalian host [3]. Nonetheless, the largest
39 biodiversity of this protist family is among trypanosomatids that usually carry out their entire life
40 cycle inside insects [4-6]. *Herpetomonas muscarum ingenoplastis* was isolated and described by Rogers
41 & Wallace in 1971 [7]. This parasite was capable of infecting flies from nine different genera, being
42 the most prevalent *Phormia*. In artificial infections, it demonstrates high host specificity towards
43 *Phormia regina* [7], which is a Palearctic fly found in North America and Northern Europe. Also
44 known as 'black blow fly', it plays an important role in the ecosystem via carrion decomposition and
45 nutrient recycling [8].

46 A BLAST analysis of the single available sequence of *H. muscarum ingenoplastis* (18S rRNA gene,
47 GenBank Acc. number KX901631) revealed that it does not cluster with any other member of the
48 genus *Herpetomonas*. Instead, its closest phylogenetic relatives (Trypanosomatidae spp. MCC-01,
49 MCC-02, MCC-03, GMO-05, D44-1, G42, PNG60, and MCZ-14) form a separate group on the
50 phylogenetic tree of trypanosomatids [9-11]. Here, we sequenced the whole genome of *H. muscarum*
51 *ingenoplastis* combining MinION and Illumina.

52 2. Results

53 The Illumina sequencing yielded 100,372,731 reads, out of which 89.61% presented a Phred Q
54 score of 30 or higher, and a mean quality score of 37.55. Regarding the MinION sequencing, the
55 starting DNA presented a good quality with a DNA Integrity Number (DIN) of 9.1. After shearing,
56 the majority of DNA (90% of the total) was composed of fragments from 3,208 bp to 46,456 bp, with
57 an average size of 10,112 bp. Subsequently, a 1D sequencing library was run for approximately 43 h
58 in a flow cell, generating a total of 2,402,163 reads. After base-calling, 88% of the total reads passed
59 the mean quality score threshold of 7. The read N50 for those that passed the filter was 6,514, with
60 2637 reads longer than 20 kb, whereas the longest read was 54.8 kb.

61 The assembly generated using Canu consisted of 340 contigs, which were polished by the
62 Illumina data using PILON. It resulted in a genome size of 35.09 Mb with an N50 of 375,483 bp, G+C
63 content of 53.73%. The average coverages were 428X (MinION) and 270X (Illumina). The draft
64 genome was aligned to *H. muscarum* reference genome by LastZ (v. 1.04.00) revealing that only 1.5%
65 of the latter is covered by the assembly with an identity 80% or higher. This result underscores
66 previous data from our research group, which indicate that this isolate is phylogenetically distant
67 from all described trypanosomatids and must therefore be assigned to a new genus [11]. The
68 automated annotation revealed a total of 8,619 genes. The long reads generated by the third-
69 generation sequencing technologies, such as Oxford Nanopore Technologies (ONT), are particularly
70 suitable to address the challenges associated with trypanosomatids genome, allowing direct
71 determination of the full sequence of large clusters of repetitive sequences without collapsing them.
72 In the fast changing field of long-read DNA sequencing, the Fiocruz Protist Collection decided to
73 provide full genomic sequences of reference strains, as a strategic decision to boost science and
74 promote Culture Collections [12].

75 3. Materials and Methods

76 *H. muscarum ingenoplastis* is cryopreserved at Fiocruz Protist Culture Collection (COLPROT)
77 (<http://colprot.fiocruz.br>), voucher number COLPROT021. This specimen is also available at the
78 American Type Culture Collection (ATCC30259). Flagellates were grown in a biphasic medium
79 NNN/LIT (Novy-MacNeal-Nicolle/Liver Infusion Tryptose) supplemented with 10% fetal bovine
80 serum. The genomic DNA was extracted using PureLink Genomic DNA mini kit (Invitrogen) from
81 cells in the late logarithmic phase of growth. DNA quality control was performed by measuring the
82 absorbance at 260/230, concentration was determined using Qubit, and DNA integrity was analyzed
83 by 0.8% agarose gel electrophoresis and using an Agilent 2200 TapeStation system with the Genomic
84 DNA Screen Tape assay. Genome sequencing was performed using Illumina TruSeq DNA PCR-Free
85 kit on Illumina HiSeq 4000 platform with 2 × 100 paired-end reads. Sequence quality metrics were
86 assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

87 The long reads were obtained using the ONT MinION sequencer on FLO-MIN106 R9v flow cells.
88 We prepared the library using the 1D Genomic DNA by ligation (SQK-LSK108) protocol. Briefly, high
89 molecular weight DNA (1.3 µg) was sheared with a g-TUBE (Covaris) to an average fragment length
90 of 8 Kb. The sheared DNA was repaired using the FFPE Repair mix (New England Biolabs), polished
91 and an A overhang was added with NEBNext End Prep Module (New England Biolabs).
92 Subsequently, adapters (Adapter Mix AMX1D) were ligated using the Blunt/TA Ligase Master Mix
93 (New England Biolabs). Between each step, DNA was cleaned using Ampure XP beads (Beckman
94 Coulter) in a 1:1 proportion. The final library was loaded on the MinION flow cell and monitored by
95 MinKNOW software (version 1.15.1) during a 48-h sequencing time. Generated reads were base-

96 called in real time and assembled using Canu v1.4 [13]. The assembly was corrected using PILON
97 [14]. The final generated assembly was assessed by QUAST (Quality Assessment Tool for Genome
98 Assemblies) [15] in Icarus genome browser [16]. The Companion webtool
99 (<https://companion.sanger.ac.uk/>) was used for gene prediction and annotation, and *Leishmania major*
100 as a reference genome [17].

101

102 **Author Contributions:** Conceptualization, C.M.dL., B.B., V.Y., P.B., P.G., J.G., M.V.; methodology, C.M.dL., B.B.,
103 J.A., R.H., A.B., K.A.M., H.L.C., P.B.; data analysis, C.M.dL., B.B., J.A., R.H., A.B., K.A.M., H.L.C., P.B., P.G., J.G.,
104 M.V.; resources, C.M.dL., P.G., J.G., M.V.; data curation, C.M.dL., B.B., J.A., R.H., A.B., V.Y., K.A.M., H.L.C., P.B.,
105 P.G., J.G., M.V.; writing—original draft preparation, C.D.dL.; writing—review and editing, C.M.dL., B.B., J.A.,
106 R.H., A.B., V.Y., K.A.M., H.L.C., P.B., P.G., J.G., M.V.; supervision, M.V.; project administration, C.M.dL.;
107 funding acquisition, C.M.dL., V.Y., P.G., J.G., M.V.

108 **Funding:** This research was funded in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -
109 Brasil (CAPES) - Finance Code 001, CAPES-COFECUB (Comité Français d'Évaluation de la Coopération
110 Universitaire et Scientifique avec le Brésil, program 923/18), European Regional Funds (project "Centre for
111 Research of Pathogenicity and Virulence of Parasites" CZ.02.1.01/0.0/0.0/16_019/0000759), Conselho Nacional de
112 Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo a Pesquisa do Estado do Rio de Janeiro
113 (FAPERJ) and Fundação Oswaldo Cruz (Fiocruz). P.B. is a Scientific Logistics Manager of the Genomics Platform
114 of University of Louvain. C.M.dL. was awarded a scholarship from the CAPES Foundation (PVEX -
115 88881.169912/2018-01), in order to conduct part of his Research as a Visiting scholar at Université de Louvain
116 (Brussels/Belgium).

117 **Acknowledgments:** The correspondence author is particularly grateful for COBIO from CNPq, the scientific and
118 career experience were deeply boosted by the extra challenges granted by this funding agency. The authors
119 thank the National Lottery (Belgium) and the Foundation against Cancer (2010-101, Belgium) for their support
120 to the Genomics Platform of University of Louvain and de Duve Institute, as well as the Fonds de la Recherche
121 Scientifique - FNRS Equipment Grant UN035.17 for the «Big data analysis cluster for NGS at UCL».

122 **Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the
123 study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to
124 publish the results.

125 Appendix A

126 Data access

127 This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the
128 accession VFSE00000000. The version described in this paper is version VFSE01000000.

129 References

- 130 1. Lukeš, J.; Butenko, A.; Hashimi, H.; Maslov, D.A.; Votýpka, J.; Yurchenko, V. Trypanosomatids are much
131 more than just trypanosomes: clues from the expanded family tree. *Trends Parasitol* **2019**, *34*(6), 466-480.
132 doi: 10.1016/j.pt.2018.03.002.
- 133 2. Nussbaum, K.; Honek, J.; Cadmus, C.M.; Efferth, T. Trypanosomatid parasites causing neglected diseases.
134 *Curr Med Chem* **2010**, *17*(15), 1594-617. doi: 10.2174/092986710790979953.
- 135 3. Sangenito, L.S.; da Silva Santos, V.; d'Avila-Levy, C.M.; Branquinha, M.H.; Santos, A.L.S.; Oliveira, S.S.C.
136 Leishmaniasis and Chagas Disease - Neglected Tropical Diseases: Treatment Updates. *Curr Top Med Chem*
137 **2019**, *19*(3), 174-177. doi: 10.2174/156802661903190328155136.
- 138 4. Podlipaev, S.A. Insect trypanosomatids: the need to know more. *Mem Inst Oswaldo Cruz* **2000**, *95*(4), 517-
139 22. doi: 10.1590/s0074-02762000000400013
- 140 5. Maslov, D.A.; Votýpka, J.; Yurchenko, V.; Lukeš, J. Diversity and phylogeny of insect trypanosomatids: all
141 that is hidden shall be revealed. *Trends Parasitol* **2013**, *29*(1), 43-52. doi: 10.1016/j.pt.2012.11.00
- 142 6. d'Avila-Levy, C.M.; Boucinha, C.; Kostygov, A.; Santos, H.L.; Morelli, K.A.; Grybchuk-Ieremenko, A.;
143 Duval, L.; Votýpka, J.; Yurchenko, V.; Grellier, P.; Lukeš, J. Exploring the environmental diversity of
144 kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem Inst Oswaldo Cruz* **2015**, *110*(8),
145 956-965. doi: 10.1590/0074-02760150253

- 146 7. Rogers, W.E.; Wallace, F.G. Two new subspecies of *Herpetomonus muscurum* (Leidy, 1856) Kent, 1880. *J*
147 *Protozool* **1971**, *18*(4), 645-649. doi: 10.1111/j.1550-7408.1971.tb03390.x
- 148 8. Putnam, R.J. The role of carrion-frequenting arthropods in the decay process. *Ecol Entomol* **1978**, *3*(2), 133–
149 139. doi: 10.1111/j.1365-2311.1978.tb00911.x
- 150 9. Votýpka, J.; Klepetková, H.; Jirků, M.; Kment, P.; Lukeš, J. Phylogenetic relationships of trypanosomatids
151 parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. *Int J Parasitol* **2012**, *42*(5), 489-500. doi:
152 10.1016/j.ijpara.2012.03.007.
- 153 10. Týč, J.; Votýpka, J.; Klepetková, H.; Suláková, H.; Jirků, M.; Lukeš, J. Growing diversity of trypanosomatid
154 parasites of flies (Diptera: Brachycera): frequent cosmopolitanism and moderate host specificity. *Mol*
155 *Phylogenet Evol* **2013**, *69*(1), 255-264. doi: 10.1016/j.ympev.2013.05.024
- 156 11. Boucinha, C.; Rodrigues, A.C.; Santos, H.L.C.; Healers, R.; Vikkula, M.; Branquinha, M.H.; Santos, A.L.S.;
157 Grellier, P.; Morelli, K.A.; d'Avila-Levy, C. M. An update and historical overview of trypanosomatids
158 available at Fundação Oswaldo Cruz Protist Collection (Fiocruz/COLPROT). *Mem Inst Oswaldo Cruz* (under
159 review).
- 160 12. d'Avila-Levy, C.M.; Yurchenko, V.; Votýpka, J.; Grellier, P. Protist Collections: Essential for Future
161 Research. *Trends Parasitol* **2016**, *32*(11), 840-842. doi: 10.1016/j.pt.2016.08.001
- 162 13. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: scalable and accurate
163 long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **2017**, *27*(5), 722-736.
164 doi: 10.1101/gr.215087.116.
- 165 14. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.;
166 Wortman, J.; Young, S.K.; Earl, A.M. Pilon: an integrated tool for comprehensive microbial variant
167 detection and genome assembly improvement. *PLoS One* **2014**, *9*(11):e112963. doi:
168 10.1371/journal.pone.0112963.
- 169 15. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: quality assessment tool for genome assemblies.
170 *Bioinformatics* **2013**, *29*(8), 1072-1075. doi: 10.1093/bioinformatics/btt086.
- 171 16. Mikheenko, A.; Valin, G.; Prjibelski, A.; Saveliev, V.; Gurevich, A. Icarus: visualizer for de novo assembly
172 evaluation. *Bioinformatics* **2016**, *32*(21), 3321-3323. doi: 10.1093/bioinformatics/btw379
- 173 17. Steinbiss, S.; Silva-Franco, F.; Brunk, B.; Foth, B.; Hertz-Fowler, C.; Berriman, M.; Otto, T.D. Companion: a
174 web server for annotation and analysis of parasite genomes. *Nucleic Acids Res* **2016**, *44*(W1):W29-34. doi:
175 10.1093/nar/gkw292