



Brief Report

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

Claudia M. d'Avila-Levy ^{1,2*}, Bertrand Bearzatto ³, Jérôme Ambroise ³, Raphaël Helaers ², Anzhelika Butenko ^{4,5}, Vyacheslav Yurchenko ^{5,6}, Karina A. Morelli ^{1,7}, Helena L. C. Santos ¹, Pascal Brouillard ², Philippe Grellier ⁸, Jean-Luc Gala ³ and Miikka Vikkula ²

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

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

³ Centre de Technologies Moléculaires Appliquées, Université Catholique de Louvain, B-1200 Brussels, Belgium; jean-luc.gala@uclouvain.be

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

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

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

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

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

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

Received: 18 November 2019; Accepted: 10 February 2020; Published: 13 February 2020

Abstract: Here, we present first draft genome sequence of the trypanosomatid *Herpetomonas muscarum ingenoplastis*. This parasite was isolated repeatedly in the black blowfly, *Phormia regina*, and it forms a phylogenetically distinct clade in the Trypanosomatidae family.

Keywords: genome assembly; monoxenous trypanosomatids; insect trypanosomatids; Trypanosomatidae; whole genome

1. Introduction

The family Trypanosomatidae encompasses parasites of vertebrates, invertebrates, or plants [1]. Chagas disease, leishmaniasis, and human African trypanosomiasis are human diseases caused by *Trypanosoma cruzi*, *Leishmania* spp. and *Trypanosoma brucei* sensu lato, respectively [2]. These parasites affect about 22 million people worldwide and alternate their life cycle between an insect vector and a mammalian host [3]. Therefore, the research is concentrated in these disease-inflicting parasites, however, the largest biodiversity of this family is among trypanosomatids that usually infects insects as the single host [4–6]. *Herpetomonas muscarum ingenoplastis* was isolated and described by Rogers and Wallace in 1971 [7]. This parasite was capable of infecting flies from nine different genera, with *Phormia* being the most prevalent genus. In artificial infections, it demonstrates high host specificity towards *Phormia regina* [7], which is a Palearctic fly found in North America and Northern Europe, also known as “black blow fly” which plays a key role in the ecosystem via carrion decomposition and nutrient recycling [8].

A BLAST analysis of the single available sequence of *H. muscarum ingenoplastis* (18S rRNA gene, GenBank Acc. number KX901631) revealed that it does not cluster with any other member of the genus *Herpetomonas*. Instead, its closest phylogenetic relatives (Trypanosomatidae spp. MCC-01, MCC-02, MCC-03, GMO-05, D44-1, G42, PNG60, and MCZ-14) form a separate group on the phylogenetic tree of trypanosomatids [9,10]. Here, we sequenced the whole genome of *H. muscarum ingenoplastis* combining MinION and Illumina.

2. Results and Discussion

The Illumina sequencing yielded 100,372,731 reads, out of which 89.61% presented a Phred Q score of 30 or higher, and a mean quality score of 37.55. Regarding the MinION sequencing, the starting DNA presented a good quality with a DNA Integrity Number (DIN) of 9.1. After shearing, the majority of DNA (90% of the total) was composed of fragments from 3208 bp to 46,456 bp, with an average size of 10,112 bp. Subsequently, a one-dimensional (1D) sequencing library was run for approximately 43 h in a flow cell, generating a total of 2,402,163 reads. After basecalling, 88% of the total reads passed the mean quality score threshold of 7. The ones that passed the filter had a N50 of 6514, with 2637 reads longer than 20 kb, whereas the longest read was 54.8 kb.

The assembly generated using the MinION reads in Canu consisted of 340 contigs, which were polished by the Illumina data using PILON (Appendix A). It resulted in a genome size of 35.09 Mb with an N50 of 375,483 bp, and G + C content of 53.73%. The average coverages were 428X (MinION) and 270X (Illumina). The automated annotation revealed a total of 8619 genes (Table S1 in Supplementary Materials), including putative mitochondrial proteins. The draft genome was aligned to *H. muscarum* reference genome (GCA_000482205.1) by LastZ (v. 1.04.00) revealing that only 1.5% of the latter presented an identity of 80% or higher with the draft genome. The analysis of the gGAPDH gene, widely used in barcoding and taxonomic studies [6], revealed an identity of 85% over 713 nucleotides between *H. muscarum ingenoplastis* and *H. muscarum*. The maximum likelihood (ML) and Bayesian inference (BI) phylogenetic trees reconstructed with gGAPDH were generally in agreement with the described phylogeny of the group [11] (Figure 1) and indicated that this isolate is phylogenetically distant from all described trypanosomatids, and therefore must be assigned to a new genus, as previously suggested [9,10].

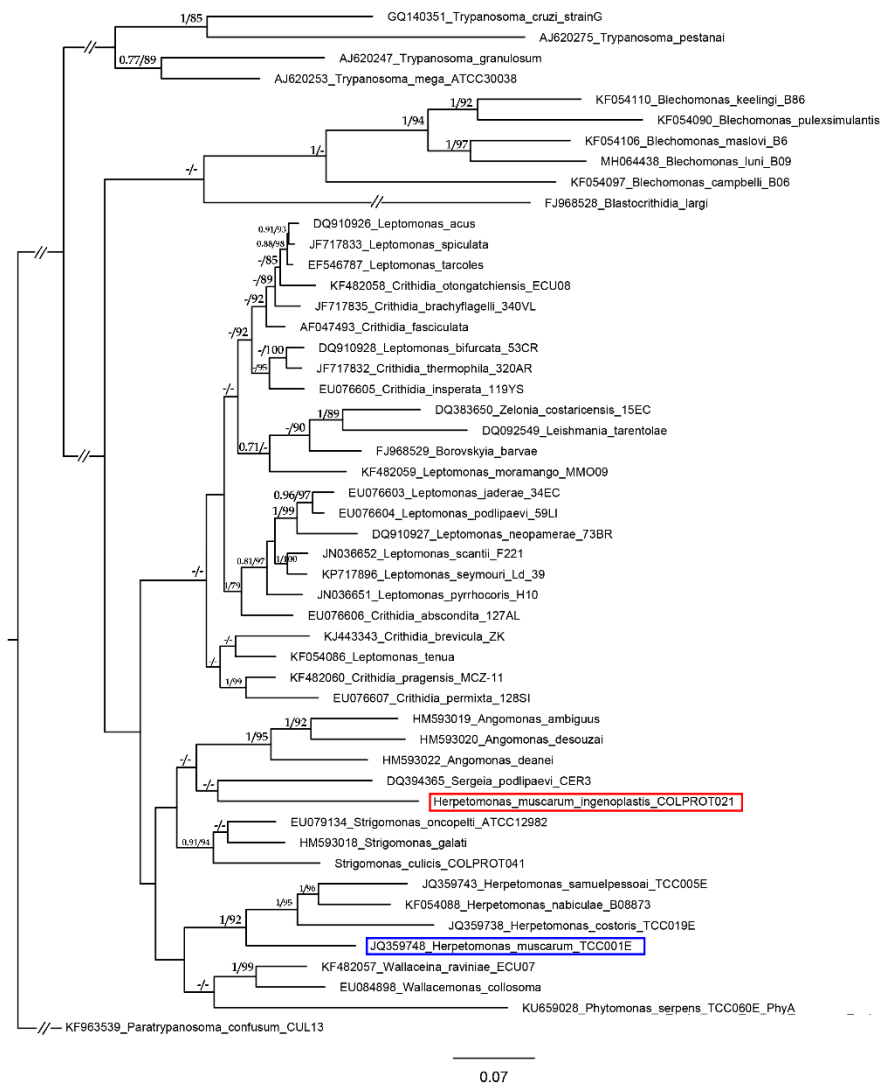


Figure 1. Phylogenetic analysis by ML and BI of *H. muscarum ingenoplastis*. The tree is based on the partial sequences of gGAPDH from COLPROT021 and GenBank sequences. The numbers at the top of each node denote Bayesian posterior probability and maximum likelihood bootstrap values. Dashes (-) indicate bootstrap support below 70% or different topology. The tree was rooted with the sequences from *Paratrypanosoma confusum*. Double-crossed branches are at 50% of their original lengths. The scale bar denotes the number of substitutions per site.

The genomes in public database are unevenly distributed among the Trypanosomatidae family and the vast majority are concentrated in *Leishmania* and *Trypanosoma* genera (more than 50). There are five *Crithidia* spp. genomes, which are mainly used as models for biochemical, molecular, and cellular biology phenomena. There are five genomes available from representatives of the Strigomonadinae subfamily, which has attracted attention from researchers due to the possibility of deepening the understanding on endosymbiosis [12]. There are three genomes from *Phytomonas* spp. that have driven research because of the phytopathogenicity of some species [13]. Then, among the formally described genera of the family (more than 20), there are two *Leptomonas* spp. genomes, and one genome for *Paratrypanosoma*, *Endotrypanum*, *Blechomonas*, *Lotmaria*, and *Herpetomonas*. Therefore, expanding the diversity of representatives of the family with whole genome sequences would help to elucidate the phylogeny, unveil hidden biodiversity, and pinpoint specific features of the genomes and cell biology of poorly studied taxa. Particularly, *H. muscarum ingenoplastis* attracted our attention due to old reports on its exquisite cell biology, that is, the presence of double-flagellate promastigotes [7]. In the fast-changing field of long-read DNA sequencing, the Fiocruz Protist Collection decided to

provide full genomic sequences of reference strains, as a strategic decision to boost science and promote Culture Collections [14].

3. Materials and Methods

H. muscarum ingenoplastis is cryopreserved at Fiocruz Protist Culture Collection (COLPROT) (<http://colprot.fiocruz.br>), voucher number COLPROT-021. This specimen is also available at the American Type Culture Collection (ATCC 30259). Flagellates were grown in a biphasic medium NNN/LIT (Novy-MacNeal-Nicolle/Liver Infusion Tryptose) supplemented with 10% fetal bovine serum. The genomic DNA was extracted using PureLink Genomic DNA mini kit (Invitrogen) from cells in the late logarithmic phase of growth. DNA quality control was performed by measuring the absorbance at 260/230, concentration was determined using Qubit, and DNA integrity was analyzed by 0.8% agarose gel electrophoresis and using an Agilent 2200 TapeStation system with the Genomic DNA Screen Tape assay. Genome sequencing was performed using Illumina TruSeq DNA PCR-Free kit on Illumina HiSeq 4000 platform with 2 × 100 paired-end reads. Sequence quality metrics were assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

The long reads were obtained using the ONT MinION sequencer on FLO-MIN106 R9v flow cells. We prepared the library using the 1D Genomic DNA by ligation (SQK-LSK108) protocol. Briefly, high molecular weight DNA (1.3 µg) was sheared with a g-TUBE (Covaris) to an average fragment length of 8 Kb. The sheared DNA was repaired using the FFPE Repair mix (New England Biolabs), polished and an A overhang was added with NEBNext End Prep Module (New England Biolabs). Subsequently, adapters (Adapter Mix AMX1D) were ligated using the Blunt/TA Ligase Master Mix (New England Biolabs). Between each step, DNA was cleaned using Ampure XP beads (Beckman Coulter) in a 1:1 proportion. The final library was loaded on the MinION flow cell and monitored by MinKNOW software (version 1.15.1) during a 48 h sequencing time. The generated reads were basecalled, in real time, and assembled using Canu v1.4 [15]. The assembly was corrected with the Illumina data using Pilon [16]. The final generated assembly was assessed by QUAST (quality assessment tool for genome assemblies) [17] in Icarus genome browser [18]. The Companion webtool (<https://companion.sanger.ac.uk/>) was used for gene prediction and annotation, and *Leishmania major* as a reference genome [19]. For the phylogenetic inference, gGAPDH was PCR-amplified from gDNA, sequenced, and deposited in GenBank under the accession number KX901490.1, as described elsewhere [20]. Subsequently, gGAPDH sequences were aligned using multiple sequence alignment with high accuracy and high throughput (MAFFT) online server and manually refined in BioEdit [21]. To identify the phylogenetic position of the isolate, phylogenetic trees were created using *Paratrypanosoma confusum*, as the outgroup [11]. Phylogenetic trees were constructed using two probabilistic methods, ML and BI, which were based on GTR + G substitution model, according to the Akaike Information Criterion and Bayesian information criterion (BIC) using Jmodeltest [22]. The ML tree was created in METAPIGA v2.0 [23] and BI in MrBayes v3.2 [24]. By the analysis of 1000 replicates and the MCMC algorithm, with four chains, the bootstrap values were determined for the ML and BI, respectively. For each 1000 generations, chains were sampled out of a total of 10⁷ generations. Convergence was evaluated by the mean standard deviation of split frequencies that were lower than the recommended values (<0.01). For each dataset, the first quarter of the selected trees was excluded as burn-in, and the nodal support and consensus tree topology were assessed from the remaining samples as posterior probability values.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: Putative protein list based on the automated annotation by Companion, using *Leishmania major* as reference.

Author Contributions: Conceptualization, C.M.dL., B.B., V.Y., P.B., P.G., J.G., and M.V.; methodology, C.M.dL., B.B., J.A., R.H., A.B., K.A.M., H.L.C., and 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., and M.V.; resources, C.M.dL., P.G., J.G., and 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., P.G., J.G., and M.V.; writing—original draft preparation, C.M.dL.; writing—review and editing, C.M.dL., B.B., J.A., R.H., A.B., V.Y., K.A.M., H.L.C., P.B., P.G., J.G., and M.V.; supervision, M.V.; project administration, C.M.dL.; funding acquisition, C.M.dL., V.Y., P.G., J.G., and M.V.

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

Acknowledgments: The correspondence author is particularly grateful to COBIO from CNPq, the scientific and career experience were deeply boosted by the extra challenges granted by this funding agency. The authors thank the National Lottery (Belgium) and the Foundation against Cancer (2010-101, Belgium) for their support to the Genomics Platform of University of Louvain and de Duve Institute, as well as the Fonds de la Recherche Scientifique - FNRS Equipment Grant U.N035.17 for the «Big data analysis cluster for NGS at UCL».

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

Appendix A

Data access

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VFSE00000000. The version described in this paper is version VFSE00000000.1.

References

1. Lukeš, J.; Butenko, A.; Hashimi, H.; Maslov, D.A.; Votýpka, J.; Yurchenko, V. Trypanosomatids are much more than just trypanosomes: Clues from the expanded family tree. *Trends Parasitol.* **2019**, *34*, 466–480, doi:10.1016/j.pt.2018.03.002.
2. Nussbaum, K.; Honek, J.; Cadmus, C.M.; Efferth, T. Trypanosomatid parasites causing neglected diseases. *Curr. Med. Chem.* **2010**, *17*, 1594–1617, doi:10.2174/092986710790979953.
3. Sengenito, L.S.; da Silva Santos, V.; d'Avila-Levy, C.M.; Branquinha, M.H.; Santos, A.L.S.; Oliveira, S.S.C. Leishmaniasis and Chagas Disease—Neglected Tropical Diseases: Treatment Updates. *Curr. Top. Med. Chem.* **2019**, *19*, 174–177, doi:10.2174/156802661903190328155136.
4. Podlipaev, S.A. Insect trypanosomatids: The need to know more. *Mem. Inst. Oswaldo Cruz* **2000**, *95*, 517–522, doi:10.1590/s0074-02762000000400013.
5. Maslov, D.A.; Votýpka, J.; Yurchenko, V.; Lukeš, J. Diversity and phylogeny of insect trypanosomatids: All that is hidden shall be revealed. *Trends Parasitol.* **2013**, *29*, 43–52, doi:10.1016/j.pt.2012.11.00.
6. d'Avila-Levy, C.M.; Boucinha, C.; Kostygov, A.; Santos, H.L.; Morelli, K.A.; Grybchuk-Ieremenko, A.; Duval, L.; Votýpka, J.; Yurchenko, V.; Grellier, P.; et al. Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem. Inst. Oswaldo Cruz* **2015**, *110*, 956–965, doi:10.1590/0074-02760150253.
7. Rogers, W.E.; Wallace, F.G. Two new subspecies of *Herpetomonus muscurum* (Leidy, 1856) Kent, 1880. *J. Protozool.* **1971**, *18*, 645–649, doi:10.1111/j.1550-7408.1971.tb03390.x.
8. Putnam, R.J. The role of carrion-frequenting arthropods in the decay process. *Ecol. Entomol.* **1978**, *3*, 133–139, doi:10.1111/j.1365-2311.1978.tb00911.x.
9. Votýpka, J.; Klepetková, H.; Jirků, M.; Kment, P.; Lukeš, J. Phylogenetic relationships of trypanosomatids parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. *Int. J. Parasitol.* **2012**, *42*, 489–500, doi:10.1016/j.ijpara.2012.03.007.
10. Týč, J.; Votýpka, J.; Klepetková, H.; Suláková, H.; Jirků, M.; Lukeš, J. Growing diversity of trypanosomatid parasites of flies (Diptera: Brachycera): Frequent cosmopolitanism and moderate host specificity. *Mol. Phylogenet. Evol.* **2013**, *69*, 255–264, doi:10.1016/j.ympev.2013.05.024.

11. Ishemgulova, A.; Butenko, A.; Kortiřová, L.; Boucinha, C.; Grybchuk-Ieremenko, A.; Morelli, K.A.; Tesařová, M.; Kraeva, N.; Grybchuk, D.; Pánek, T.; et al. Molecular mechanisms of thermal resistance of the insect trypanosomatid *Crithidia thermophila*. *PLoS ONE* **2017**, *12*, e0174165.
12. Brunoro, G.V.F.; Menna-Barreto, R.F.S.; Garcia-Gomes, A.S.; Boucinha, C.; Lima, D.B.; Carvalho, P.C.; Teixeira-Ferreira, A.; Trugilho, M.R.O.; Perales, J.; Schwämmle, V.; et al. Quantitative Proteomic Map of the Trypanosomatid *Strigomonas culicis*: The Biological Contribution of its Endosymbiotic Bacterium. *Protist* **2019**, *170*, 125698.
13. Camargo, E.P. *Phytomonas* and other trypanosomatid parasites of plants and fruit. *Adv. Parasitol.* **1999**, *42*, 29–112.
14. d'Avila-Levy, C.M.; Yurchenko, V.; Votýpka, J.; Grellier, P. Protist Collections: Essential for Future Research. *Trends Parasitol.* **2016**, *32*, 840–842, doi:10.1016/j.pt.2016.08.001.
15. Koren, S.; Walenz, B.P.; Berlin, K.; Miller, J.R.; Bergman, N.H.; Phillippy, A.M. Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **2017**, *27*, 722–736, doi:10.1101/gr.215087.116.
16. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963, doi:10.1371/journal.pone.0112963.
17. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075, doi:10.1093/bioinformatics/btt086.
18. Mikheenko, A.; Valin, G.; Prjibelski, A.; Saveliev, V.; Gurevich, A. Icarus: Visualizer for de novo assembly evaluation. *Bioinformatics* **2016**, *32*, 3321–3323, doi:10.1093/bioinformatics/btw379.
19. Steinbiss, S.; Silva-Franco, F.; Brunk, B.; Foth, B.; Hertz-Fowler, C.; Berriman, M.; Otto, T.D. Companion: A web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* **2016**, *44*, W29–W34, doi:10.1093/nar/gkw292.
20. Hall, T.A. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
21. Posada, D.; Buckley, T.R. Model selection and model averaging in phylogenetics: Advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Syst. Biol.* **2004**, *53*, 793–808.
22. Helaers, R.; Milinkovitch, M.C. MetaPIGA v2.0: Maximum likelihood large phylogeny estimation using the metapopulation genetic algorithm and other stochastic heuristics. *BMC Bioinform.* **2010**, *11*, 379.
23. Huelsenbeck, J.P.; Ronquist, F.; Nielsen, R.; Bollback, J.P. Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **2001**, *294*, 2310–2314.
24. Ronquist, F.; Teslenko, M.; van der Mark, P.; Ayres, D.L.; Darling, A.; Hohna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **2012**, *61*, 539–542.

