Protistology

SHORT COMMUNICATION

Transcriptome of *Lafontella* sp., an obscure relative of *Phytomonas* and *Herpetomonas*

Amanda T. S. Albanaz[1,†], Anzhelika Butenko[1,2,3,†]*,
Alexei Yu. Kostygov[1,4], Jan Votýpka[2,5] and
Vyacheslav Yurchenko[1]

[1] *Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czechia;*
[2] *Institute of Parasitology, Czech Academy of Sciences, 370 05 České Budějovice, Czechia;*
[3] *University of South Bohemia, Faculty of Sciences, 370 05 České Budějovice, Czechia;*
[4] *Zoological Institute of the Russian Academy of Sciences, 199034 St. Petersburg, Russia;*
[5] *Department of Parasitology, Faculty of Science, Charles University, 128 44 Prague, Czechia*

[†] Shared first authorship

Summary

The trypanosomatid subfamily Herpetomonadinae includes plant-pathogenic dixenous genus *Phytomonas* as well as two sister monoxenous genera − *Herpetomonas* and *Lafontella*. The latter one is very poorly studied with the only information on its biology being its affinity to muscid and calliforid flies and the inability to survive at elevated temperature. To shed more light on this obscure genus and get insights into the evolution of gene content in the subfamily Herpetomonadinae, we conducted transcriptome sequencing of *Lafontella* sp. GMO-01 and comparative analysis of its predicted arsenal of proteins with those of other trypanosomatids. Our results show considerable similarity between the two monoxenous genera of Herpetomonadinae in the total number of proteins, exceeding that in the majority of other trypanosomatids as well as content of functional orthogroups. Using *Lafontella* and *Herpetomonas* as the closest outgroups allowed inferring the genes uniquely lost by *Phytomonas*, their dixenous cousin. The fact that GC content of *Lafontella* is intermediate between that of *Herpetomonas* and *Phytomonas* correlating with the phylogenetic distances separating the three genera suggests that a more thorough analysis of this flagellate can shed new light on the origin of dixeny in this subfamily.

Key words: Trypanosomatidae, Herpetomonadinae, dixeny, orthogroups

* **Corresponding author**: Anzhelika Butenko, Institute of Parasitology CAS, Branišovská 1160/31, 370 05 České Budějovice, Czechia; anzhelika.butenko@paru.cas.cz

Introduction

The parasitic flagellates of the family Trypanosomatidae are considered one of the best-studied groups of protists, but the bulk of research has been focused on human pathogens of the genera *Trypanosoma* and *Leishmania* (Kostygov et al., 2021). Nevertheless, in the last two decades, previously neglected monoxenous trypanosomatids, i. e. those, whose development is restricted to a single (usually, an insect) host, started to attract substantially more attention as convenient models for studies in various biological fields and as interesting objects *per se* (Maslov et al., 2019).

Out of 24 currently recognized trypanosomatid genera 19 are monoxenous and five are dixenous (switching between a vertebrate or plant host and an arthropod or leech vector) (Kostygov et al., 2021). They are united into subfamilies, of which only two – Leishmaniinae and Herpetomonadinae contain both monoxenous and dixenous genera and, therefore, are interesting for the investigation of the origin of dixeny (Jirků et al., 2012; Yurchenko et al., 2016). Of these, the members of Leishmaniinae, comprising human pathogens and being more diverse with its nine described genera (Kostygov and Yurchenko, 2017), have received significantly more attention. However, the subfamily Herpetomonadinae undoubtedly deserves scrutiny, since it includes the only plant-parasitic genus *Phytomonas* as well as two monoxenous ones: *Herpetomonas*, a speciose genus with a high potential for adaption to new hosts, and poorly known *Lafontella* (Yurchenko et al., 2021).

The genus *Lafontella* has been established recently and so far contains only a single described species – *L. mariadeanei* (Yurchenko et al., 2016). Previously, this species was classified in *Herpetomonas* (Borghesan et al., 2013), but significant molecular divergence as well as presence of unique cells – elongated endomastigotes with a long coiled intracellular flagellum (Yoshida et al., 1978) – justified its separation into a separate sister genus (Yurchenko et al., 2016). This species was originally isolated from the gut of the false stable fly, *Muscina stabulans* (Muscidae) in Brazil, but later phylogenetically close trypanosomatids were detected in the blowflies *Chrysomya megacephala* from Ghana, *Calliphora vomitoria* from Bulgaria and *Chrysomya putoria* from Madagascar (all Calliphoridae) (Týč et al., 2013; Votýpka et al., 2020).

Nothing is known about the biology of *Lafontella mariadeanei* except for its host affinity and the inability to withstand elevated (37 °C) temperature (Yoshida et al., 1978). To shed more light on this obscure genus and get insights into the evolution of gene content in the subfamily Herpetomonadinae, we conducted transcriptome sequencing of *Lafontella* sp. GMO-01, an isolate, which is very close to the type strain ATCC 30708 of *Lafontella mariadeanei*. Our preliminary analysis showed considerable similarity between the two monoxenous genera of Herpetomonadinae and, by using these as the closest outgroups, allowed to infer the genes lost in their dixenous cousin, *Phytomonas*.

Material and methods

Culture, RNA isolation and transcriptome sequencing

Lafontella sp. GMO-01 was isolated in 2009 from *Chrysomya megacephala* (Diptera) in Abrafo, Ghana (5°19′53″N, 1°22′49″E) and cultivated at 23 °C in Schneider's Drosophila medium (Merck, St. Louis, USA) supplemented with 10% fetal bovine serum (FBS) (Thermo Fisher Scientific, Waltham, USA), 100 μg/ml of streptomycin and 100 Units/ml of penicillin (both from Merck). RNA isolation was performed using the RNeasy minikit (Qiagen, Hilden, Germany) following the manufacturer's protocol. TrueSeq stranded mRNA library was prepared and sequenced using Illumina NovaSeq 6000 instrument (Macrogen, Seoul, Korea), producing approximately 32 million 151 nt long paired-end reads (Suppl. Table S1). Raw sequencing reads were deposited in the NCBI database under BioProject accession PRJNA949447.

Transcriptome assembly

Raw Illumina sequencing reads were adapter- and quality-trimmed using Fastp v.0.20.1 (Chen et al., 2018), and only paired-end reads with the length exceeding 50 nt were retained for further analysis. Read quality and adapter content were evaluated before and after trimming using FastQC v.0.11.9 (Andrews, 2010). The trimmed reads were error corrected by Karect, with the correction accuracy assessed on a preliminary assembly with '-align' and '-eval' settings (Allam et al., 2015).

De novo transcriptome assembly was performed using rnaSPAdes v3.13.0 (Bushmanova et al., 2019), and only transcripts of 200 bp and longer were retained for further analysis. The assembly was checked for potential contamination with

BlobTools v.1.1.1 (Laetsch and Blaxter, 2017). Scaffolds showing high-quality (i.e., nucleotide sequence identity >95% and query coverage >85%) BLASTN v.2.13.0 (Camacho et al., 2009) hits to non-euglenozoan sequences in NCBI nucleotide database (download date: 2022-5-8) were discarded as potential contaminants.

Transcript redundancy was reduced by running CD-HIT v.8.1 in the CD-HIT-EST mode with 90% sequence identity threshold (Li and Godzik, 2006). The resulting transcripts were gap-filled using GapCloser v.1.12 module from SOAPde-novo2 (Luo et al., 2012). Putative coding regions were predicted using TransDecoder v.5.5.0 (https://github.com/TransDecoder/TransDecoder).

The transcriptome assembly was deposited NCBI under the accession number GKLZ00000000.

ASSEMBLY QUALITY ASSESSMENT AND FUNCTIONAL ANNOTATION

Standard assembly metrics, such as length, number of transcripts, etc., were calculated with QUAST v.5.0.2 and SeqKit v0.16.0 (Shen et al., 2016; Mikheenko et al., 2018). Assembly completeness was assessed using two different metrics: i) proportion of the so-called universal single copy orthologs from Euglenozoa_odb10 and Eukaryota_odb10 reference databases as estimated by BUSCO v.5 (Manni et al., 2021) and ii) percentage of transcriptomic reads mapping back to the assembly. For the latter, the reads were mapped back to the transcriptome using Bwa-mem2 v.2.2.1 (Vasimuddin et al., 2019) and sorted with SAMtools v.1.16.1 (Danecek et al., 2021). The mapping rate was assessed with BamTools v.2.4.1 (Barnett et al., 2011).

Transcriptome-derived proteins were functionally annotated based on the homologs identified with BLASTP (e-value threshold of 1e-10) in the genomes of the following trypanosomatids: *Leishmania major* Friedlin (TriTrypDB-62), *Herpetomonas muscarum* TCC001E (GCA_000482205.1), *Phytomonas* sp. Hart1 (GCA_000982615.1), *Trypanosoma brucei* TREU927 (GCA_000002445.1), and *Paratrypanosoma confusum* CUL13 (TriTrypDB-62). Proteins with no match were submitted to a second homology-based search restricted to the whole family Trypanosomatidae (taxid:5654) in the NCBI non-redundant protein database with the same e-value threshold. Additionally, we submitted all predicted proteins of *Lafontella* sp., *H. muscarum, P. francai, P. serpens, Phytomonas* spp. EM1 and

Hart1, *L. major* Friedlin, *T. brucei* TREU927 and *P. confusum* to eggNOG-mapper v.2.0.1 (Cantalapiedra et al., 2021) with eggNOG 5 as a reference database (Huerta-Cepas et al., 2019). Eukaryota were chosen as the target taxon, and the transfer of both experimental and electronic annotations was allowed. Assigned categories of clusters of orthologous groups (COGs) were visualized using R package ggplot2 v.3.4.2 (Wickham, 2009). Proteins with no orthologs in the eggNOG database were added to the category 'no COG assigned'.

ORTHOLOGY INFERENCE

Groups of orthologous proteins (OGs) were inferred using OrthoFinder v.2.5.5 (Emms and Kelly, 2019) with default settings and a reference dataset comprising *Lafontella* sp. GMO-01, *Leishmania major* Friedlin (TriTrypDB-62), *Herpetomonas muscarum* TCC001E (GCA_000482205.1), *Phytomonas francai* TCC-064 (GCA_001766655.1), *Phytomonas serpens* 9T (GCA_000331125.1), *Phytomonas* sp. EM1 (GCA_000582765.1) and Hart1 (GCA_000982615.1), *Trypanosoma brucei* TREU927 (GCA_000002445.1), and *Paratrypanosoma confusum* CUL13 (TriTrypDB-62). Shared and species-specific OGs were visualized using UpSetR software (Lex et al., 2014).

Results and discussion

TRANSCRIPTOME ASSEMBLY

Transcriptome sequencing resulted in the generation of 31.7 million reads out of which 30.5 million successfully passed adapter- and quality-trimming process (Suppl. Table S1). The obtained assembly had a total length of 24.0 Mb with ~190× read coverage. Our decontamination procedure filtered out short sequences with predominantly low read coverage that appeared to be of Chordata origin (Suppl. Fig. S1). The final decontaminated non-redundant assembly comprises 13,455 assembled transcripts with 11,786 predicted proteins (Table 1; Suppl. Table S1). The assembly has a high degree of completeness as judged by presence of all BUSCOs from the Euglenozoa database and 99.8% rate of read mapping (Table 1; Suppl. Table S1).

The total number of encoded proteins in *Lafontella* sp. GMO-01 and its closest studied relative, *H. muscarum*, is similar: both species encode ~ 12,000 proteins (Table 1) (Sloan et al., 2019). This is about

Table 1. Comparison of gene content and features between the transcriptomic assembly of *Lafontella* sp. and genomic assemblies of other trypanosomatids.

	Total number of proteins	Median length of CDS, nt	Average GC content in CDSs, %	BUSCO scores according to euglenozoa_odb10 database, %				
				Complete	Single-copy	Duplicated	Fragmented	Missing
Lafontella sp.	11,786	1,149	57.53	100	99.2	0.8	0.0	0.0
Herpetomonas muscarum	12,162	1,314	65.86	92.3	88.5	3.8	4.6	3.1
Phytomonas francai	5,946	1,316	55.32	92.3	91.5	0.8	5.4	2.3
Phytomonas serpens	6,183	987	52.33	76.2	75.4	0.8	12.3	11.5
Phytomonas sp. EM1	6,381	1,107	52.57	90.0	90.0	0.0	9.2	0.8
Phytomonas sp. Hart1	6,451	1,023	51.00	80.8	80.8	0.0	18.5	0.7
Paratrypanosoma confusum	8,659	1,392	62.27	96.2	96.2	0.0	2.3	1.5
Leishmania major	8,424	1,431	62.54	100.0	99.2	0.8	0.0	0.0
Trypanosoma brucei	8,784	1,158	50.93	97.7	97.7	0.0	1.5	0.8

twice as much as in the plant-infecting *Phytomonas* spp., whose streamlined genomes encode only ~6,000 proteins (Table 1) and 30-40% more than in the two human pathogens taken here as references − *Leishmania major* and *Trypanosoma brucei*. Considering that monoxenous trypanosomatids have simpler life cycles compared to their dixenous relatives, such a result is counterintuitive. It should be noted that comparing proteins predicted in transcriptomes and genomes may result in under- or overestimation of gene counts. However, underestimating the gene counts due to lack of expression of inducible genes can be neglected here, since in trypanosomatids protein-coding genes are polycistronically transcribed (Maslov et al., 2019). As for the potential overestimation, it was minimized here by transcript redundancy removal.

The median length of coding sequences (CDS) in *Lafontella* sp. GMO-01 is 1,149 nt, which is about 12% smaller than that in *H. muscarum*, but falls within the range of CDS lengths for representatives of the subfamily Herpetomonadinae (Table 1). Interestingly, the GC content in the coding sequences of *Lafontella* sp. (57.5%) is intermediate between that of *H. muscarum* (65.9%) and those of *Phytomonas* spp. (51-55%), but obviously closer to the latter (Table 1), which correlates with the phylogenetic distances between the three genera on the tree. It is worth noting that there is a difference of approximately 15% in the GC content of coding regions between the representative of

Herpetomonadinae with the highest and the lowest values observed in *H. muscarum* and *Phytomonas* sp. Hart1, respectively. This difference reflects the intra and intergenomic variation observed in organisms both within and outside the family Trypanosomatidae (Romiguier and Roux, 2017).

FUNCTIONAL DIVERSITY OF TRANSCRIPTOME-DERIVED PROTEINS

To gain insights into the functional diversity of proteins of *Lafontella* sp. GMO-01 and compare it to other Herpetomonadinae and trypanosomatids outside of this subfamily, we clustered the proteins within each genome and transcriptome according to the eggNOG nomenclature (Cantalapiedra et al., 2021) (Fig. 1). Monoxenous representatives of Herpetomonadinae exhibit remarkably similar COG profiles, with the largest functional categories being "posttranslational modification, protein turnover, chaperones," "signal transduction mechanisms," "translation, ribosomal structure and biogenesis," as well as "amino acid transport and metabolism" (Fig. 1). These two species primarily differ in the category of proteins involved in translation, where *Lafontella* sp. possesses 1.5% fewer proteins (Fig. 1). Conversely, this species contains approximately 1.3% more proteins involved in cell wall/membrane/envelope biogenesis than *H. muscarum*. Importantly, the COG profiles of the two trypanosomatids exhibit more similarity to that
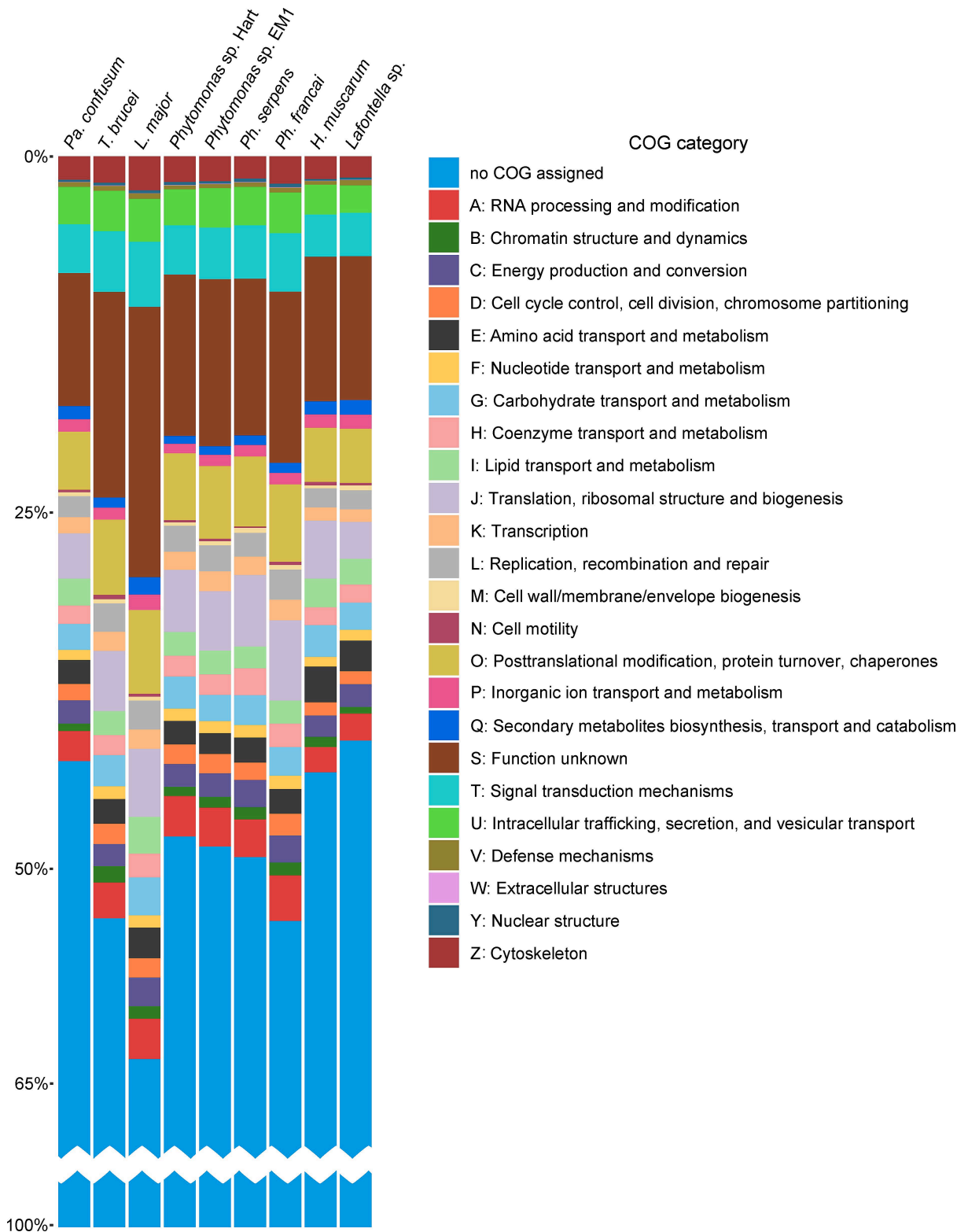
Fig. 1. Functional classification of trypanosomatid proteins according to eggNOG nomenclature. The percentage on the X-axis corresponds to the proportion of proteins falling into each category for each species indicated on the Y-axis.

Fig. 2. UpSet plot showing orthologous groups sharing among nine trypanosomatid species, including *Lafontella* sp. GMO-01. Only 50 largest intersections are depicted. Species-specific genes were also considered. The number of orthologous groups is shown on the Y-axis and above the bar plot; the X-axis shows the species composition for each intersection.

of the monoxenous mosquito parasite *P. confusum* than to those of dixenous parasites (Fig. 1). This correlation suggests that the genetic landscape of trypanosomatid parasites is to a large extent shaped by their lifestyle and host specificity.

Out of approximately 75,000 proteins used for OG inference, around 70,000 (93%) were assigned to 8,725 groups containing at least two proteins (Supp. Table S2). Of these, 3,323 OGs were shared by all analyzed species (Fig. 2) and likely represented housekeeping genes inherited from the last common trypanosomatid ancestor (LCTA). The set of 461 OGs, absent only in *Phytomonas* spp., contained proteins apparently present in LCTA but dispensable for the survival of the flagellates from the plant-pathogenic genus (Fig. 2). The functional analysis of these OGs revealed heme-dependent proteins and those included into heme-dependent protein complexes: cytochrome c oxidase subunits, the respective assembly factors, and heme-dependent peroxidases (Suppl. Table S3). In addition, *Phytomonas* spp. residing in carbohydrate-rich plant tissues lost multiple genes encoding amino acid transporters, peptidases, and aminotransferases (Suppl. Table S3). These results are in line with previous inferences made for the genomes of this genus (Kořený et al., 2012; Porcel et al., 2014; Butler et al., 2017)

Among the 11,786 transcriptome-derived proteins of *Lafontella* sp. GMO-01, approximately 11,210 (~95%) were clustered into 7,445 OGs, of which 1,357 (~12%) were assigned to 267 species-specific groups containing at least two proteins (Supp. Table S2). Furthermore, 567 groups contained proteins encoded by single-copy species-specific genes. Within the 147 OGs uniquely shared between *Lafontella* sp. and *H. muscarum* (Fig. 2), there are proteins annotated as amastins and peptidases, likely involved in interactions with dipteran hosts (Suppl. Table S4). The two species also share genes encoding transposable elements, although to assess their prevalence more precisely a further analysis is required using the genome sequence of *Lafontella* sp.

Acknowledgements

References

Allam A., Kalnis P. and Solovyev V. 2015. Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. Bioinformatics. 31 (21): 3421−3428. https://doi.org/10.1093/bioinformatics/btv415

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Barnett D.W., Garrison E.K., Quinlan A.R., Stromberg M.P. and Marth G.T. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics. 27 (12): 1691−1692. https://doi.org/10.1093/bioinformatics/btr174

Borghesan T.C., Ferreira R.C., Takata C.S., Campaner M. et al. 2013. Molecular phylogenetic redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a genus of insect parasites associated with flies. Protist. 164 (1): 129−152. https://doi.org/10.1016/j.protis.2012.06.001

Bushmanova E., Antipov D., Lapidus A. and Prjibelski A.D. 2019. rnaSPAdes: A *de novo* transcriptome assembler and its application to RNA-seq data. Gigascience. 8 (9): giz100. https://doi.org/10.1093/gigascience/giz100

Butler C.E., Jaskowska E. and Kelly S. 2017. Genome sequence of *Phytomonas françai*, a cassava (*Manihot esculenta*) latex parasite. Genome Announc. 5 (2): e01266-01216. https://doi.org/10.1128/genomeA.01266-16

Camacho C., Coulouris G., Avagyan V., Ma N. et al. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10: 421. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra C.P., Hernández-Plaza A., Letunic I., Bork P. and Huerta-Cepas J. 2021. egg NOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. 38 (12): 5825−5829. https://doi.org/10.1093/molbev/msab293

Chen S., Zhou Y., Chen Y. and Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinform. 34 (17): i884−i890. https://doi.org/10.1093/bioinformatics/bty560

Danecek P., Bonfield J.K., Liddle J., Marshall J. et al. 2021. Twelve years of SAMtools and BCFtools. Gigascience. 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008

Emms D.M. and Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative

genomics. Genome. Biol. 20 (1): 238. https://doi.org/10.1186/s13059-019-1832-y

Huerta-Cepas J., Szklarczyk D., Heller D., Hernandez-Plaza A. et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47 (D1): D309−D314. https://doi.org/10.1093/nar/gky1085

Jirků M., Yurchenko V.Y., Lukeš J. and Maslov D.A. 2012. New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. J. Eukaryot. Microbiol. 59 (6): 537−547. https://doi.org/10.1111/j.1550-7408.2012.00636.x

Kořený L., Sobotka R., Kovářová J., Gnipová A. et al. 2012. Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. Proc Natl Acad Sci U S A. 109 (10): 3808-3813. https://doi.org/10.1073/pnas.1201089109

Kostygov A.Y., Karnkowska A., Votýpka J., Tashyreva D. et al. 2021. Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. Open Biol. 11 (3): 200407. https://doi.org/10.1098/rsob.200407

Kostygov A.Y. and Yurchenko V. 2017. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). Folia Parasitol 64: 020. https://doi.org/10.14411/fp.2017.020

Laetsch D.R. and Blaxter M.L. 2017. BlobTools: interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Research. 6: 1287. https://doi.org/10.12688/f1000research.12232.1

Lex A., Gehlenborg N., Strobelt H., Vuillemot R. and Pfister H. 2014. UpSet: visualization of intersecting sets. IEEE Trans Vis Comput Graph. 20 (12): 1983−1992. https://doi.org/10.1109/TVCG.2014.2346248

Li W. and Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 22 (13): 1658−1659. https://doi.org/10.1093/bioinformatics/btl158

Luo R., Liu B., Xie Y., Li Z. et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience. 1 (1): 18. https://doi.org/10.1186/2047-217X-1-18

Manni M., Berkeley M.R., Seppey M., Simro F.A. and Zdobnov E.M. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol.

Biol. Evol. 38 (10): 4647−4654. https://doi.org/10.1093/molbev/msab199

Maslov D.A., Opperdoes F.R., Kostygov A.Y., Hashimi H. et al. 2019. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. Parasitology. 146 (1): 1−27. https://doi.org/10.1017/S0031182018000951

Mikheenko A., Prjibelski A., Saveliev V., Antipov D. and Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 34 (13): i142−i150. https://doi.org/10.1093/bioinformatics/bty266

Porcel B.M., Denoeud F., Opperdoes F.R., Noel B. et al. 2014. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. PLoS. Genet. 10 (2): e1004007. https://doi.org/10.1371/journal.pgen.1004007

Romiguier J. and Roux C. 2017. Analytical biases associated with GC-content in molecular evolution. Front Genet. 8: 16. https://doi.org/10.3389/fgene.2017.00016

Shen W., Le S., Li Y. and Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE. 11 (10): e0163962. https://doi.org/10.1371/journal.pone.0163962

Sloan M.A., Brooks K., Otto T.D., Sanders M.J. et al. 2019. Transcriptional and genomic parallels between the monoxenous parasite *Herpetomonas muscarum* and *Leishmania*. PLoS. Genet. 15 (11): e1008452. https://doi.org/10.1371/journal.pgen.1008452

Týč J., Votýpka J., Klepetková H., Šuláková H. et al. 2013. Growing diversity of trypanosomatid parasites of flies (Diptera: Brachcera): frequent cosmopolitism and moderate host specificity. Mol. Phylogenet. Evol. 69 (1): 255−264. https://doi.org/10.1016/j.ympev.2013.05.024

Vasimuddin M., Misra S., Li H. and Aluru S., 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems, 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS 2019). IEEE, Rio de Janeiro, Brazil, pp. 314−324.

Votýpka J., Kment P., Yurchenko V. and Lukeš J. 2020. Endangered monoxenous trypanosomatid parasites: a lesson from island biogeography. Biodivers. Conserv. 29 (13): 3635−3667. https://doi.org/10.1007/s10531-020-02041-2

Wickham H., 2009. ggplot2: Elegant graphics for data analysis. Springer, New York.

Yoshida N., Freymuller E. and Wallace F.G. 1978. *Herpetomonas mariadeanei* sp. n. (Protozoa, Trypanosomatidae) from *Muscina stabulans* (Fallen, 1816) (Diptera, Muscidae). J. Protozool. 25 (4): 421–425. https://doi.org/10.1111/j.1550-7408.1978.tb04161.x

Yurchenko V., Butenko A. and Kostygov A.Y. 2021. Genomics of Trypanosomatidae: where we stand and what needs to be done? Pathogens. 10 (9): 1124. https://doi.org/10.3390/pathogens10091124

Yurchenko V., Kostygov A., Havlová J., Grybchuk-Ieremenko A. et al. 2016. Diversity of trypanosomatids in cockroaches and the description of Herpetomonas tarakana sp. n. J. Eukaryot. Microbiol. 63 (2): 198–209. https://doi.org/10.1111/jeu.12268

Supplementary material

Fig. S1. Blobplots for the transcriptome assembly before and after decontamination.

Table S1. Transcriptome sequencing and assembly statistics for *Lafontella* sp.

Table S2. Sizes of protein families as inferred by the analysis of orthologous groups.

Table S3. Functional annotation of OGs lost in *Phytomonas* spp.

Table S4. Functional annotation of OGs shared only by *Lafontella* sp. and *H. muscarum*.