







# Somy evolution in the honey bee infecting trypanosomatid parasite *Lotmaria passim*

Lindsey M. Markowitz <sup>1,2</sup> Anthony Nearman,<sup>1</sup> Zexuan Zhao <sup>2</sup> Dawn Boncristiani,<sup>1</sup> Anzhelika Butenko,<sup>3,4,5</sup> Luis Miguel de Pablos,<sup>6,7</sup> Arturo Marin <sup>8</sup> Guang Xu <sup>9</sup> Carlos A. Machado <sup>2</sup> Ryan S. Schwarz,<sup>10</sup> Evan C. Palmer-Young,<sup>1,\*</sup> Jay D. Evans <sup>1,\*</sup>

<sup>1</sup>USDA-ARS Bee Research Laboratory, 10300 Baltimore Ave, BARC-East Bldg. 306 Rm 313, Beltsville, MD 20705, USA

<sup>2</sup>Department of Biology, University of Maryland, Biology-Psychology Building, 4094 Campus Drive, College Park, MD 20742, USA

<sup>3</sup>Czech Academy of Sciences, Institute of Parasitology, České Budějovice 370 05, Czech Republic

<sup>4</sup>Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava 710 00, Czech Republic

<sup>5</sup>Faculty of Science, University of South Bohemia, České Budějovice 370 05, Czech Republic

<sup>6</sup>Department of Parasitology, Biochemical and Molecular Parasitology Group CTS-183, University of Granada, Granada 18071, Spain

<sup>7</sup>Institute of Biotechnology, University of Granada, Granada 18071, Spain

<sup>8</sup>Omics Bioinformatics S.L., Calle Senderos 2, Bajo, Granada 18005, Spain

<sup>9</sup>Department of Microbiology, University of Massachusetts, Fernald Hall, Amherst MA 01003, USA

<sup>10</sup>Department of Biology, Fort Lewis College, 1000 Rim Drive, Durango, CO 81301, USA

\*Corresponding author: Jay D. Evans, USDA-ARS Bee Research Laboratory, 10300 Baltimore Ave, BARC-East Bldg. 306 Rm 313, Beltsville, MD 20705, USA.

Email: jay.evans@usda.gov; Evan C. Palmer-Young, USDA-ARS Bee Research Laboratory, 10300 Baltimore Ave, BARC-East Bldg. 306 Rm 313, Beltsville, MD 20705, USA.

Email: evan.palmer-young@usda.gov

*Lotmaria passim* is a ubiquitous trypanosomatid parasite of honey bees nestled within the medically important subfamily Leishmaniinae. Although this parasite is associated with honey bee colony losses, the original draft genome—which was completed before its differentiation from the closely related *Crithidia mellificae*—has remained the reference for this species despite lacking improvements from newer methodologies. Here, we report the updated sequencing, assembly, and annotation of the BRL-type (Bee Research Laboratory) strain (ATCC PRA-422) of *Lotmaria passim*. The nuclear genome assembly has been resolved into 31 complete chromosomes and is paired with an assembled kinetoplast genome consisting of a maxicircle and 30 minicircle sequences. The assembly spans 33.7 Mb and contains very little repetitive content, from which our annotation of both the nuclear assembly and kinetoplast predicted 10,288 protein-coding genes. Analyses of the assembly revealed evidence of a recent chromosomal duplication event within chromosomes 5 and 6 and provided evidence for a high level of aneuploidy in this species, mirroring the genomic flexibility employed by other trypanosomatids as a means of adaptation to different environments. This high-quality reference can therefore provide insights into adaptations of trypanosomatids to the thermally regulated, acidic, and phytochemically rich honey bee hindgut niche, which offers parallels to the challenges faced by other Leishmaniinae during the challenges they undergo within insect vectors, during infection of mammals, and exposure to antiparasitic drugs throughout their multi-host life cycles. This reference will also facilitate investigations of strain-specific genomic polymorphisms, their role in pathogenicity, and the development of treatments for pollinator infection.

**Keywords:** *Lotmaria passim* strain BRL ; ATCC PRA-422; Leishmaniinae; Trypanosomatidae; trypanosomatid; PacBio; Hi-C; monoxenous; polyploidy; aneuploidy

## Introduction

The family Trypanosomatidae (Euglenozoa: Kinetoplastea) is a diverse group of unicellular flagellated parasites that infect plants and many animals, including certain groups of insects (Maslov *et al.* 2019). Trypanosomatid genomics has focused on insect-vectored, dixenous (two host) taxa in the genera *Trypanosoma* and *Leishmania* due to their clinical relevance as parasites (Albanaz *et al.* 2023; Kostygov *et al.* 2024). Consequently, these genera have been the main focus of genomic efforts and have been key to understanding the phylogenetic history, geographic radiation, population dynamics, and therapeutic drug resistance in this family (Lukeš *et al.* 2018). In contrast, the solely insect-infecting monoxenous trypanosomatids are biodiverse examples of obligate endosymbionts that have received less attention

(Lukeš *et al.* 2018). However, a recent surge in work on these monoxenous relatives has offered several genomic insights into the metabolism, host-specific adaptations for infection, and intraspecific diversity of these species (Kostygov *et al.* 2024) while also offering clues to the evolution of host shifts and parasitism of humans and other mammals (Flegontov *et al.* 2016).

Bumblebee (*Bombus* spp.) and honey bee (*Apis* spp.) pollinators are insects in which monoxenous trypanosomatids seem to thrive (McGhee and Cosgrove 1980; Schwarz *et al.* 2015) as parasites (see, for example: Schmid-Hempel 2001; Schwarz *et al.* 2015; Koch *et al.* 2017; Gómez-Moracho *et al.* 2020; MacInnis *et al.* 2023). Thus, bees are significant and readily accessible models for trypanosomatid-host coevolution and for insight into the way these parasites affect individual fitness and colony success. High-quality genomes of the bumblebee trypanosomatids

*Crithidia bombi* and *Crithidia expoeki* have aided in this effort and provide a basis for studies on population genomics and the identification of genes under selection (Schmid-Hempel et al. 2018). The honey bee parasite *Lotmaria passim* was first distinguished from its close relative, *Crithidia mellificae*, in 2015 (Schwarz et al. 2015) and has since been recognized as the dominant trypanosomatid infecting honey bees worldwide (Morimoto et al. 2013; Ravoet et al. 2013; 2015; Arismendi et al. 2016; Vejnovic et al. 2018; Xu et al. 2018; Castelli et al. 2019).

The first genome assembly of *Lotmaria passim* was published before its differentiation from *Crithidia mellificae* (Runckel et al. 2014) and has remained the reference for this species despite lacking improvements from newer methodologies and the association of *Lotmaria passim* with honey bee colony loss and reduced individual bee survival in some experiments (Cornman et al. 2012; Gómez-Moracho et al. 2020). Although the absence of a high-quality reference genome has not prevented recent investigations into *Lotmaria passim* physiology and cell biology (Buendia-Abad et al. 2022; Carreira de Paula et al. 2024), transcriptomics (Liu et al. 2020), proteomics and the role specific genes play during infection (Yuan et al. 2024), it does preclude more detailed analysis of important features such as ploidy, repetitive sequences, and an overall gene arrangement.

Here, we report the sequencing, chromosome-level assembly, and genome annotation of the BRL-type (Bee Research Laboratory) strain of *Lotmaria passim*. Our assembly and annotation match the quality of other recently sequenced insect-specific species (Flegontov et al. 2016; Schmid-Hempel et al. 2018; Albanaz et al. 2023) and improve substantially upon the previously published draft assembly (Runckel et al. 2014). This high-quality reference could provide insights into the adaptations of trypanosomatids to the thermally regulated, acidic, and phytochemically rich honey bee hindgut niche, which offers parallels to the challenges faced by other Leishmaniinae during the challenges they undergo within insect vectors, during infection of mammals, and exposure to anti-parasitic drugs throughout their multi-host life cycles. This reference will also facilitate investigations of strain-specific genomic polymorphisms, their role in pathogenicity, and the development of treatments for pollinator infection.

## Materials and methods

### Species/Strain origin

*Lotmaria passim* BRL-type strain (ATCC PRA-422) cultures (Schwarz et al. 2015) were grown to early stationary phase in FP medium supplemented with 10% heat-inactivated fetal bovine serum (FPFB medium) (Salathé et al. 2012) at 31°C in 25-cm<sup>2</sup> vented cell culture flasks (Corning Inc., Corning, NY, USA). To generate dense cultures with sufficient DNA for long-read sequencing, ten flasks containing 12 mL of culture were combined before the library preparation steps detailed below. Cultures for sequencing of *Lotmaria passim* BRL (2024) were stored at -80°C until DNA extraction.

Culture stocks were originally isolated from the dissected ileum of an adult female honey bee at the United States Department of Agriculture Bee Research Laboratory in Beltsville, MD, and expanded in axenic cell culture before deposition as ATCC PRA-422 and cryopreservation in 2013 (Schwarz et al. 2015). Derivations obtained via axenic cell culture of this same stock were used for the *Lotmaria passim* BRL (2016) sequencing effort completed by the University of Massachusetts (GenBank Accession ID: GCA\_002216525.1). Cultures used for the *Lotmaria passim* BRL (2024) sequencing effort presented herein have since undergone shifts from antibiotic and fetal bovine serum

supplemented Insectagro DS2 medium (Schwarz et al. 2015) to antibiotic-free FPFB medium (Salathé et al. 2012) and have been grown in continuous cell culture for an estimated cumulative timeframe of six months (~500 generations), in addition to being subjected to several cryopreservation events prior to sequencing. It is likely that in response to continued passaging the once mixed strain population originating from the gut shifted to a predominantly clonally reproducing isolate with two dominant haplotypes, which is consistent with expectations for clonal reproduction of a single cell line.

## Sample preparation and sequencing

### HiFi library preparation and sequencing

High-molecular-weight genomic DNA was extracted for HiFi library preparation from pelleted cultures using the MagAttract<sup>®</sup> HMW DNA Kit (Qiagen, Hilden, Germany). Cell cultures were pelleted at 8,000 rpm for 5 minutes and washed with 1 mL of 1× PBS buffer before following all manufacturer instructions. The extracted DNA was pooled and quality-checked using a Qubit 2.0 Fluorometer (Life Technologies, CA, USA) before being stored at -20°C until shipment to the Novogene Corporation (Sacramento, CA, USA) for sequencing.

Long-read DNA sequencing was performed using 5 µg of high-molecular-weight DNA prepared as previously described to generate a PacBio Sequel II HiFi library. Analysis was completed by the Novogene Corporation (Sacramento, CA, USA) on a one-cell PacBio Sequel II (HiFi mode/cell) machine which generated 20.1 gigabases (Gb) of sequencing data with an average read length of 15,916 and a total length of 20,137,583,129 which were assembled as described below.

### Hi-C library preparation and sequencing

Genomic DNA was extracted for Hi-C library preparation from freshly pelleted cultures using the Proximo Hi-C Microbe Kit (Phase Genomics, Seattle, WA, USA). Cell cultures were pelleted at 8,000 rpm for 5 minutes and washed with 1 mL of 1X PBS buffer before following all manufacturer instructions. To obtain sufficient genetic material, 2 libraries were pooled to obtain 1,230 ng of product. A Qubit 2.0 Fluorometer (Life Technologies, CA, USA) was used to determine the concentration of each individual library (1.98 ng/µl and 27.8 ng/µl, respectively), and the resulting pooled Proximo Hi-C Library was stored at -20°C until shipment to the University of Maryland Institute for Genome Sciences for sequencing.

Short-read DNA sequencing was performed on the Proximo Hi-C Library using Illumina Nextera sequencing (S4 channel 100-bp paired-end reads). This process generated 198.5 Gb of sequencing data (2,250 million read pairs with an average length of 15,725) assembled as described below.

## Assembly

### Nuclear genome assembly

To assemble the nuclear genome remnant adapter sequences in the HiFi reads were first removed using HiFiAdapterFilt v2.0.0 (Sim et al. 2022) with the default adapter sequence database. Adapter sequences in Hi-C reads were removed using fastp v0.23.4 (Chen 2023). Sequence qualities were checked before and after filtering using fastqc v0.12.1 (Andrews et al. 2012). Haplotype-resolved contigs were assembled from HiFi and Hi-C reads using Hifiasm v0.19.6 (Cheng et al. 2021), and the two assemblies were combined and then scaffolded by Hi-C reads using YaHS v1.1 (Zhou et al. 2023). Manual correction was done in Juicebox v1.11.08 to resolve misassemblies and switch errors

(Durand et al. 2016). Finally, chromatin interactions were calculated using HiC-Pro v3.1.0 (Servant et al. 2015) and visualized using HiCPlotter v0.6.6 (Akdemir and Chin 2015).

### Kinetoplast genome assembly

The tandemly repetitive and repeat-free regions of the maxicircle were assembled separately. Reads with partial sequences similar to the cytochrome B gene (accession number: KM980180) were identified using Basic Local Alignment Search Tool (BLAST) v2.14.0 (Altschul et al. 1990). Contigs with BLAST hits were extracted and Tandem Repeats Finder v4.09 was used to locate and trim repeats (Benson 1999). A multiple sequence alignment (MSA) was built from the repeat-free contigs using MAFFT v7.525 (Katoh and Standley 2013) and a consensus sequence was built using the consensusString function from the Biostrings package v2.68.1 in R v4.3.0 (Pagès et al. 2024). All HiFi reads were aligned to the consensus sequence using Minimap2 v2.26 (Li 2018) and the alignment was sorted and indexed using SAMtools v1.6 (Danecek et al. 2021). The consensus sequence was corrected by the alignment using Pilon v1.24 (Walker et al. 2014), and all HiFi reads were aligned to the corrected repeat-free sequences again to check for homoplasmy using IGV v2.14.0 (Robinson et al. 2023).

To assemble repetitive regions, HiFi reads that were mapped to both the beginning and the end of the corrected repeat-free sequence were selected, their unmapped regions were extracted, and an MSA was built. The consensus sequence was called and then corrected by the HiFi read alignment and homoplasmy was rechecked. The corrected repetitive sequence was concatenated to the end of the repeat-free sequence, and an alignment was built again to check for misassembly and homoplasmy.

Minicircle reads were iteratively identified to reconstruct the minicircle sequences. HiFi reads were initially scanned using nhmmscan from HMMER v3.3.2 (Eddy 2011) against a customized profile Hidden Markov Model (HMM) built from the minicircle genome alignment of *Leptomonas pyrrocoris* (Gerasimov et al. 2021). The top twelve most significant reads were selected, and a self-alignment of each read was made to identify the motif of the conserved region (CR): GGGTAGGGGCGTTCACCGA. The 2 CRs and the downstream sequences were separated using AliView v1.28 (Larsson 2014) and an MSA was built, then a new profile HMM was created from the MSA and the HiFi reads were scanned again. To identify the e-value threshold, the nuclear and kinetoplast genomes were scanned, and the e-value of the top hit ( $6.30 \times 10^{-5}$ ) was selected. HiFi reads with an e-value below this threshold were used to build a new profile HMM for the next epoch and the iteration stopped when reads were no longer found. An MSA of the final minicircle read set was built, reads less than 0.1% different were grouped based on the error rates of HiFi sequencing, and consensus sequences were called to avoid duplicates.

### Repeat analysis

Repeat analysis was performed with RepeatMasker v4.1.6 (Chen 2004) using Euglenozoa as the query species and run with the Dfam database v3.8 (Storer et al. 2021).

### Annotation

Three separate gene-finding pipelines, all using an Augustus-trained model of *Leptomonas pyrrocoris* (Flegontov et al. 2016), were employed and their non-overlapping results ultimately combined. Transcript evidence for all pipelines was generated by aligning 30 separate publicly available *Lotmaria passim* short-read archives (Liu et al. 2020) using Bowtie2 v2.4.2 (very sensitive; end-to-end) (Langmead and Salzberg 2012) to our genome. The

resulting SAMtools (Li et al. 2009) converted BAM files were reformatted to GTF files using Cufflinks v2.2.1 (Trapnell et al. 2010), and then further reformatted to either an Augustus hints file or a FASTA file of mRNA transcripts using GffRead v0.12.7 (Pertea and Pertea 2020). The first gene-finding experiment was conducted using Augustus v3.5.0 (Hoff and Stanke 2019) and incorporated the *Leptomonas pyrrocoris* model and mRNA hints file. The second experiment used the Companion web server (Steinbiss et al. 2016) and incorporated the *Leptomonas pyrrocoris* model and cufflinks gtf file. The third experiment employed the Maker2 pipeline v3.01.04 (Holt and Yandell 2011) and incorporated the trained model, mRNA transcript evidence, the Euglenozoa RepBase (Bao et al. 2015) protein set for repeat masking, and the complete protein database from TriTrypDB release 67 (Shanmugasundram et al. 2023). Three rounds of Maker2 training were executed: two rounds of Snap model training and one round of Augustus.

The quality of each annotation was assessed using BUSCO 5.7.1 (Seppey et al. 2019) with Euglenozoa OrthoDB v10 proteins (Kriventseva et al. 2019) as a reference database. The first experiment's annotation produced a complete set of BUSCO proteins with no fragmentation and was then used as the basis for merging the remaining annotation experiments. We subsequently merged the non-overlapping gene predictions from the latter two experiments and assigned gene IDs using AGAT v1.4.0 (Dainat et al. 2020). A final protein set was extracted using GffRead and functional annotations were performed with InterProScan v5.66-98.0 (Jones et al. 2014) using the TIGRFAM (Haft et al. 2013) and PFAM protein family (Mistry et al. 2020) databases with residue annotation disabled, and NCBI BLAST+ (Camacho et al. 2009) using the Uniprot Database (UniProt Consortium 2015) with an e-value cut-off of  $1e^{-5}$ . The resulting functional and GO terms were added to the merged annotation using AGAT. Non-coding RNAs were annotated using cmscan Infernal 1.1.5 (Nawrocki and Eddy 2013) using Rfam-level heuristic filters and skipping all HMM filter stages. Last, we ran pseudogene predictions using Pseudo Finder v1.1.0 (Syberg-Olsen et al. 2022) against the complete protein database on TriTrypDB release 67. We retained only the predicted pseudogenes found in intergenic regions and merged those predictions with the existing annotation as previously described. Final gene annotation statistics were calculated using AGAT.

### Orthology and phylogenetics

We compared our final protein set to annotations of other Leishmaniinae: *Crithidia acanthocephali*, *Crithidia bombi*, *Crithidia expoeki*, *Crithidia brevicula*, *Crithidia fasciculata*, *Crithidia mellificae*, *Crithidia thermophilia*, *Leishmania major*, *Lotmaria passim* SF, *Leptomonas pyrrocoris*, and *Leptomonas seymouri* which originated from (Kostygov et al. 2024), in addition to *Crithidia bombi* SH and *Crithidia expoeki* SH which originated from (Schmid-Hempel et al. 2018), using OrthoFinder v2.5.5 (Emms and Kelly 2019). The results produced a total of 5,075 orthogroups with all species present and 2,833 of these consisted entirely of single-copy genes. Multisequence alignment, filtering, and trimming were performed using MAFFT v7.525 (Katoh and Standley 2013) and trimAl v1.4.rev15 (Capella-Gutiérrez et al. 2009) as described in Kostygov et al. 2024 (Kostygov et al. 2024). The resulting 986 orthogroups were subset into two separate random selections, without replacement, of 250 orthogroups. Each selection was compared using PhyloBayesMPI v1.9 (Lartillot et al. 2013) under the CATGTR model (Lartillot and Philippe 2004) over two separate chains, until the appropriate effect size (1,335) and relative difference (0.03) were achieved, at approximately 1,700 iterations per

chain. Details regarding the statistical parameters and intrinsic priors specified by the CATGTR model can be found in (Lartillot and Philippe 2004). The resulting tree file was plotted using the R packages ape v5.7-1 (Paradis and Schliep 2019) and ggtree v3.8.2 (Yu et al. 2017). No differences were found between the trees produced by separate random selections of orthogroups. The initial BUSCO analysis was repeated for all organisms listed here and plotted for comparison.

## Inference of somy

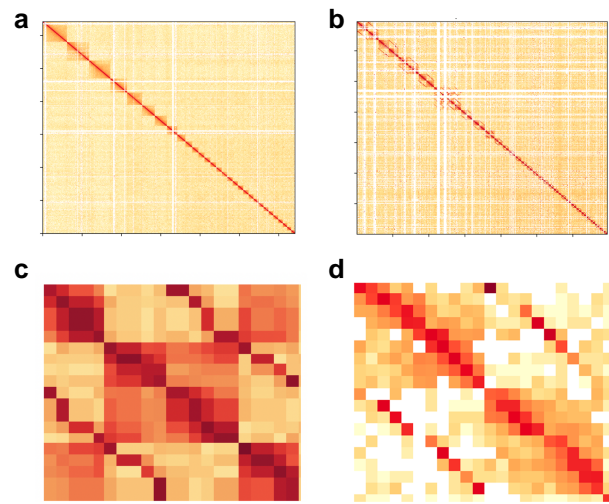
We compared data for coverage depth and variant distributions per chromosome for several *Lotmaria passim* genome sequencing efforts, including two assemblies for strain BRL: our *Lotmaria passim* BRL (2024) PacBio HiFi reads, *Lotmaria passim* strain BRL Illumina Nextseq reads from 2016 (NCBI BioProject PRJNA319530; referred to by its ATCC identifier 'PRA-422'), *Lotmaria passim* SF sequenced by Illumina Nextseq in 2016 (NCBI BioProject PRJNA319529; referred to by its ATCC identifier 'PRA-403'), and *Lotmaria passim* strains C2 and C3 isolated in 2019 from honey bees in Granada, Spain, as described by Buendía-Abad et al. (2021) and sequenced by Illumina MiSeq in 2023 (NCBI BioProject PRJNA863431, Ruiz et al., submitted).

Read mapping to our genome assembly was performed using Minimap2 (PacBio HiFi) (Li 2018) and Burrow-Wheeler Aligner (BWA)-MEM (all others) (Li 2013) with default settings. Sorting, indexing, and coverage depth of the resulting SAM files was performed with SAMtools v1.13 (Li et al. 2009). Variant calling was then performed using GATK HaplotypeCaller v4.5.0.0 (McKenna et al. 2010). Data cleaning of the resulting VCF files was performed using the R package vcfR v1.15.0 (Knaus and Grünwald 2017) as outlined by Knaus and Grünwald (2018).

Chromosome copy number was then inferred based on the three lines of evidence as generated above: median genome coverage, coverage depth by base position within each chromosome, and the distribution of allele frequencies at heterozygous sites. For depth calculations, we mapped raw genomic reads back to the assembly, calculated the median base-level coverage depth for each chromosome, and normalized depth to the median genome coverage. Normalized depths of 0.5x, 1x, 1.5x, and 2x the median were expected for monosomic, disomic, trisomic, and tetrasomic chromosomes, respectively. Patterns of coverage depth across chromosomes were visualized by graphing the base-by-base depth scores. We assessed these plots for a secondary trace at roughly half the depth of the moving average, indicative of heterozygous sites for which half of the reads for the corresponding base were attributed to a disomic chromosome's alternate haplotype. In contrast, we expected a secondary trace at roughly two-thirds that of the main trace for a trisomic chromosome, multiple secondary traces at one-fourth, one-half, and three-fourths of the main trace for a tetrasomic chromosome, and the absence of any secondary trace to suggest the absence of a distinct haplotype altogether. The distribution of allele frequencies was evaluated by inspection of variant call plots for heterozygous loci. We expected a predominantly 1:1 ratio of divergent alleles in disomic chromosomes (i.e. one allele belonging to each of two haplotypes), a 2:1 ratio for trisomic chromosomes, and a mixture of 1:1 and 3:1 frequencies in tetrasomic chromosomes (Knaus and Grünwald 2018).

## Figures

All visualizations were created using R v4.3.2 Eye Holes (R Development Core Team 2009), ggplot2 v3.5.0 (Wickham and



**Fig. 1.** Genome assembly Hi-C contact maps. Hi-C contact maps of (a) the principal nuclear genome assembly excluding chromosome 6 with 10 kb bin size, b) the diploid nuclear genome assembly with 100 kb bin size, c) haplotigs of chromosome 13 with a switch error, and (d) haplotigs of a chromosome with a manually corrected switch error. Chromosomal sequences are ordered from the top-left by scaffold size and darker colors reflect stronger linear matches. In the diploid contact map (b), alternative sequences are placed next to their corresponding principal sequences. Color intensity represents the normalized number of valid Hi-C read pairs shared between bins on the x- and y axes. Each chromosome forms a distinct block with higher within-chromosomal contact frequency. Off-diagonal lines in (b) and (c) arise from mismapping read pairs from homologous regions between haplotigs. Despite this, blocks indicating sequence boundaries remain evident. Panel C illustrates how a switch error disrupts the top-left and bottom-right blocks while Panel D illustrates the same region after manual correction.

Wickham 2016), or HiCPlotter v0.6.6 (Akdemir and Chin 2015) unless otherwise stated.

## Results and discussion

### Genome assembly and annotation

We sequenced, assembled, and annotated the first complete chromosome-level genome of *Lotmaria passim* BRL-type strain (ATCC PRA-422; herein referred to as *Lotmaria passim* BRL (2024)) leveraging the combined power of PacBio HiFi sequencing and Hi-C technologies. We achieved separation of the chromosomal haplotigs, where the longer haplotig was included in the principal assembly and the shorter was included in the alternative assembly, and the diploid assembly obtained a quality value (QV) score of 62.4367 (Fig. 1). The principal nuclear genome assembly comprises 31 chromosomal sequences (Fig. 1) with a median coverage of 161.0x, while the kinetoplast assembly includes a complete maxicircle and 30 minicircle sequences. The assembly has a total length of 33.7 Mb with a GC content of 54.5% and a contig  $N_{50}$  of 1.5 Mb (Table 1, Fig. 2). Repetitive elements identified using Euglenozoa as the query species encompass around 5.8% of the genome (1.9 Mb), and the identification of both novel repeat families and tandem repeats remains areas for further study. Our annotation of the nuclear and kinetoplast genome assemblies identified 10,288 protein-coding genes with a mean gene length of 75 bp, all of which lack introns. The protein-coding genes span 19.45 Mb (~57.7%) of the genome, and 439 pseudogenes were predicted, spanning only 320.5 Kb (less than 1% of the genome).

The nuclear genome assembly of *Lotmaria passim* BRL (2024) contains 10,270 proteins (Fig. 2), which is within the range of

**Table 1.** *Lotmaria passim* BRL 422 genome assembly and annotation summary.

Assembly	
Total genome length, Mb	33.7
GC content, %	54.5
Contig N50 <sup>a</sup> , Mb	1.5
Contig L50 <sup>b</sup>	8
Chromosome number	31
Repetitive elements (% of genome)	1,960,104 bp (5.8%)
Maxicircle assembly length, Kb	48.3
Minicircle assemblies & lengths	30 individual assemblies, ranging in length from 360 bp to 1.28 Kb
Annotation	
Gene number	10,288
Total gene length, Mb	19.45
Mean gene length, bp	1,890
Pseudogene number	439
Total pseudogene length, Kb	320.5
Mean pseudogene length, bp	730
Complete BUSCOs	130 (100%)
Complete, single-copy BUSCOs	128 (98.5%)
Complete, duplicated BUSCOs	2 (1.5%)

Mb = Megabases, Kb = kilobases, and bp = base pairs.

<sup>a</sup> Half the length of the assembly is contained within contigs of length greater than or equal to the N50.

<sup>b</sup> The L50 refers to the smallest number of contigs within which half the length of the assembly is contained.

predicted proteins (7,808 to 11,024 (Kostygov et al. 2024)) we have come to expect from other species within the subfamily Leishmaniinae and fit as expected within the Leishmaniinae clade, with *Lotmaria passim* BRL (2024) clustering together with the earlier assembly of *Lotmaria passim* strain SF (ATCC PRA-403) (Runckel et al. 2014) (Fig. 2). *Crithidia mellificae* is one exception, with 16,228 predicted proteins (Kostygov et al. 2024), a complete single-copy BUSCO score of 64.6%, and twice as many contigs as any of the other sequenced species in the clade, suggestive of a suboptimal assembly (Fig. 2).

Our assembly is larger and more complete than the earlier assembly of *Lotmaria passim* strain SF and is one of only three near-chromosome-level or chromosome-level assemblies among the 10 species and 13 individual assemblies that we compared (Fig. 2). When we analyzed the completeness of our assembly using a set of 130 Euglenozoa benchmarking universal single-copy orthologs (BUSCOs) we achieved a score of 100% (Table 1), which fits well within our expectations from closely related species (90 to 100% complete single-copy BUSCOs, Fig. 2). Our annotation subsequently identified 128 complete single-copy BUSCOs (98.5%, Table 1, Fig. 2) and 2 duplicated BUSCOs (1.5%, Table 1, Fig. 3). Both duplications were found on chromosome 5 and chromosome 6 (Fig. 3) and may be the result of a chromosomal duplication event, as discussed in detail below.

### Variation in somy within *Lotmaria passim* BRL

The expansion of genomic resources within the family Trypanosomatidae has increased exponentially in recent years, with a new focus on previously under-represented species beyond those of direct medical or veterinary importance. As more of these resources have become available, it has become abundantly clear that these genera often display varying degrees of aneuploidy

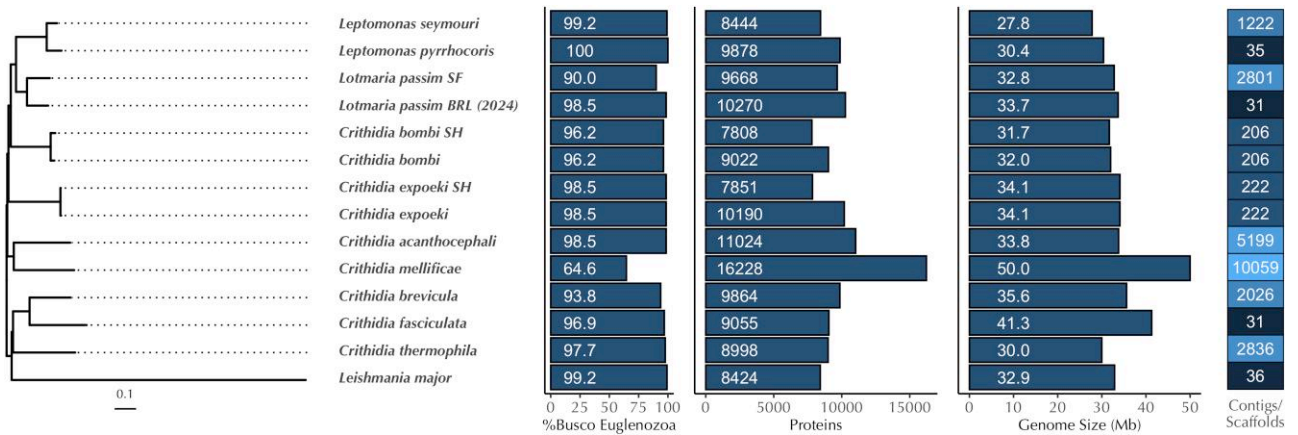
(Albanaz et al. 2023), often taking the form of either—or in some cases, both—chromosome loss or duplication. The increased plasticity observed in chromosome number and adaptive changes in somy and ploidy within Trypanosomatidae is likely one of several mechanisms to control gene expression that are employed to bypass their lack of transcription-level, gene-specific regulation (Sterkers et al. 2012; Lukeš et al. 2018; Maslov et al. 2019).

The *Lotmaria passim* BRL (2024) genome is no exception and displays several cases of aneuploidy, which we inferred from a combination of sequencing depth (Fig. 4, Supplementary Fig. 1) and variant calling (Supplementary Fig. 2). We determined that 28 of the 31 chromosomes (92% of the genome) display evidence of disomy, while only three chromosomes (chromosomes 9, 16, and 17; 8% of the genome) display evidence of trisomy (Fig. 4). Although there are no definitively tetrasomic chromosomes present in the *Lotmaria passim* BRL (2024) assembly, there is evidence that chromosomes 5 and 6 are likely derived from a single tetrasomic chromosome which then diverged from one another in an evolutionarily recent gain in chromosome number.

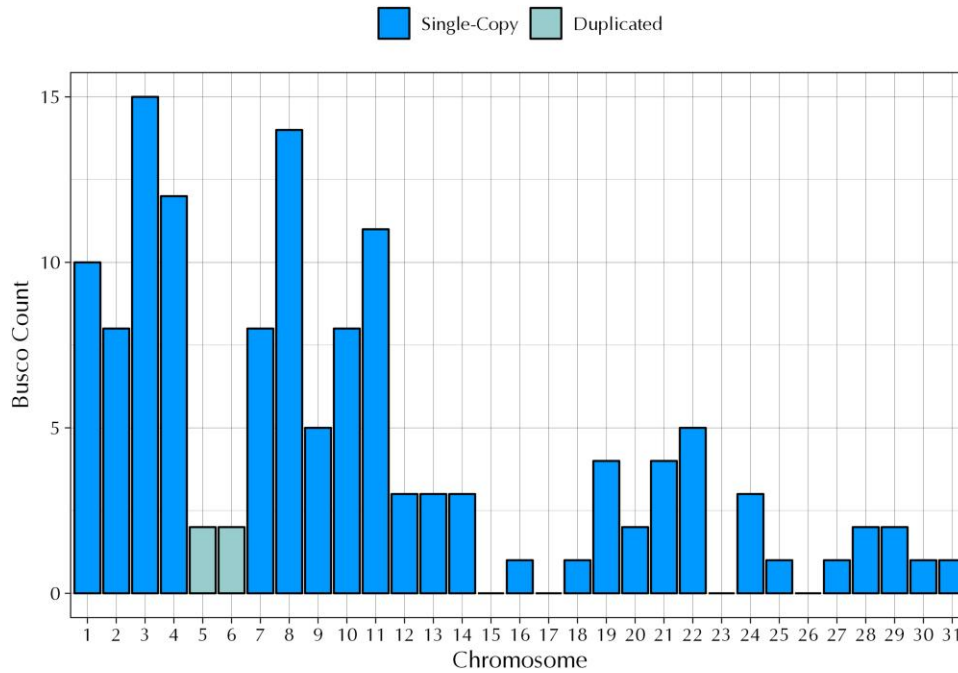
Chromosome 5 and chromosome 6 are remarkably similar (Fig. 4), and when aligned with NCBI BLAST, they have a sequence similarity score of 98.62%. Although chromosome 5 has a normalized read depth of 2.5x and chromosome 6 has a normalized read depth of 1.7x (Fig. 5), both chromosomes are likely disomic but will require further study to ascertain their somy status with more certainty. Deviations from the 2x expectation for disomic chromosomes could reflect read mapping to multiple locations due to the two regions' high degree of similarity. Although allowing for multiple read mapping may complicate inferences via normalized read depth alone, particularly for chromosome 5 where the interquartile range in depth at individual bases spans most of the 2x to 3x interval (Supplementary Fig. 1), the default parameters were left unaltered to allow for the detection of widespread patterns that may have otherwise been missed if we restricted reads to map to single sites. When we performed variant calls on both chromosomes 5 and 6 and compared our assembly against an earlier sequencing effort completed in 2016 by the University of Massachusetts (GenBank Accession ID: GCA\_002216525.1), there was a very low variant count in the BRL (2024) assembly and distributions leaning toward tetrasomy in BRL (2016) (Supplementary Fig. 2), likely due to difficulties in distinguishing between the two chromosomes.

Sequencing depth and variant calling for strain BRL (2024) did not conform to expectations for discrete levels of somy for six of the 31 nuclear chromosomes. To further demystify the observed aneuploidy of *Lotmaria passim* BRL (2024), we compared our previously calculated metrics to the normalized read depth and variant calling of the BRL (2016) assembly which was completed shortly after the strain's deposition at ATCC and lacks the approximately six months maintained in active culture (~500 generations) and several cryopreservation events undergone in our laboratory prior to sequencing of the BRL (2024) strain (Fig. 5). While this comparison revealed that the majority of the *Lotmaria passim* BRL (2024) chromosomes are indeed disomic, it also revealed several deviations from disomy.

A comparison of the two BRL assemblies identified five primary somy differences. Chromosome 16 is likely a mixture of both disomic and trisomic chromosomes, with trisomy emerging as the dominant form within strain BRL (2024), while chromosome 19 appears to be undergoing an opposite shift, with disomy evolving from trisomy (Fig. 5). These shifts are further evidenced by variant calls, which lean toward trisomy and disomy, respectively (Supplementary Fig. 2). Chromosome 23 is shifting toward disomy



**Fig. 2.** Genome statistics and phylogeny of selected species within Leishmaniinae. From left to right: Bayesian phylogenomic tree based on (~200) single-copy orthologous genes. Columns represent the percentage of complete single-copy BUSCO genes found, the number of annotated proteins, the genome size (Mb), and the number of contigs or scaffolds in species closely related to *Lotmaria passim* BRL. *Lotmaria passim* strain “SF” is based on (Runckel et al. 2014) while *Lotmaria passim* “BRL” refers to the BRL (2024) assembly presented herein. For *Crithidia bombi* and *Crithidia expoeki*, two independent annotations of the same assembly are shown, where the notation “SH” denotes the published annotations in (Schmid-Hempel et al. 2018) and the other branch corresponds to the annotations conducted in (Kostygov et al. 2024).

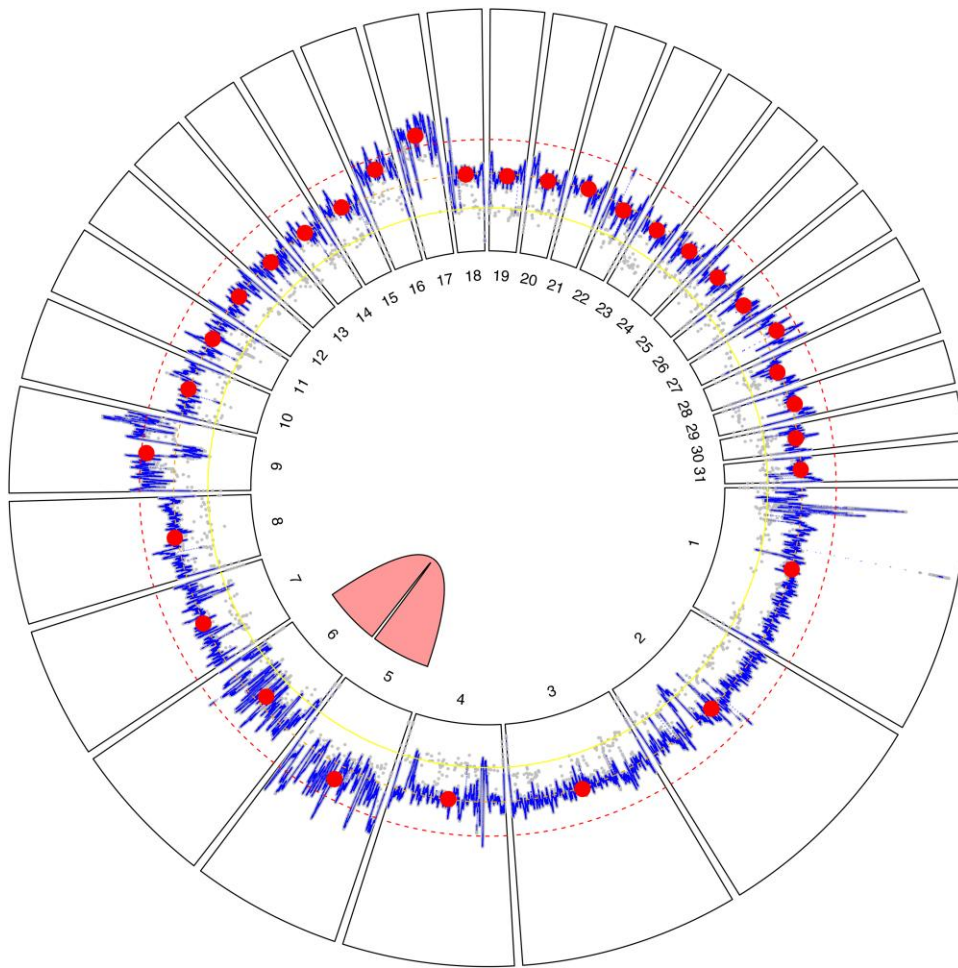


**Fig. 3.** BUSCO count by chromosome. The x-axis denotes the chromosome number and the y-axis denotes the BUSCO count derived from 130 Euklenozoa benchmarking universal single-copy orthologs (BUSCOs). Of the 31 chromosomes in the *Lotmaria passim* BRL (2024) genome, chromosomes 15, 17, 23, and 26 do not contain any BUSCOs. Chromosome 5 and chromosome 6 each contain 2 duplicated BUSCOs, an acetyl-CoA carboxylase (OrthoDB 199at33682; v10-1.orthodb.org) and peptidase M41 (OrthoDB 5001at33682; v10-1.orthodb.org).

in strain BRL (2024) and is likely derived from a mixture of haploid and diploid cells, with normalized read depth elevated from 1.2x in strain BRL (2016) to 1.8x in strain BRL (2024) (Fig. 5).

In another departure from clear-cut disomy, chromosome 27 is elevated above 2x normalized read depth (Fig. 5) at 2.5x in BRL (2016) and 2.4x in BRL (2024). Variant calls (Supplementary Fig. 2) in both strains suggest that there is a mixture of diploid and triploid cells present, likely with two identical copies and one distinct copy within the sequenced population. Although normalized read depth (Fig. 5) suggested two copies per cell of chromosome 28, the secondary coverage trace (Supplementary Fig. 1) exceeds 50% of the main

trace, suggesting a departure from the 1:1 distribution of allelic variants typical of a diploid, and variant plots (Supplementary Fig. 2) indicate close to a 2:1 ratio of alleles at heterozygous sites. This could arise if the population contains a mixture of conventional diploids, which carry two distinct copies of the chromosome, and unconventional ones that harbor two identical copies. There is strong evidence that chromosome 29 is disomic in BRL (2024) based on a doubled normalized read depth of 2.1x, although BRL (2016) leans toward trisomy at 2.5x (Fig. 5). The variant calls for each of these strains provide evidence that BRL (2016) likely contained a mixture of diploid and triploid cells at the time of sequencing, while BRL



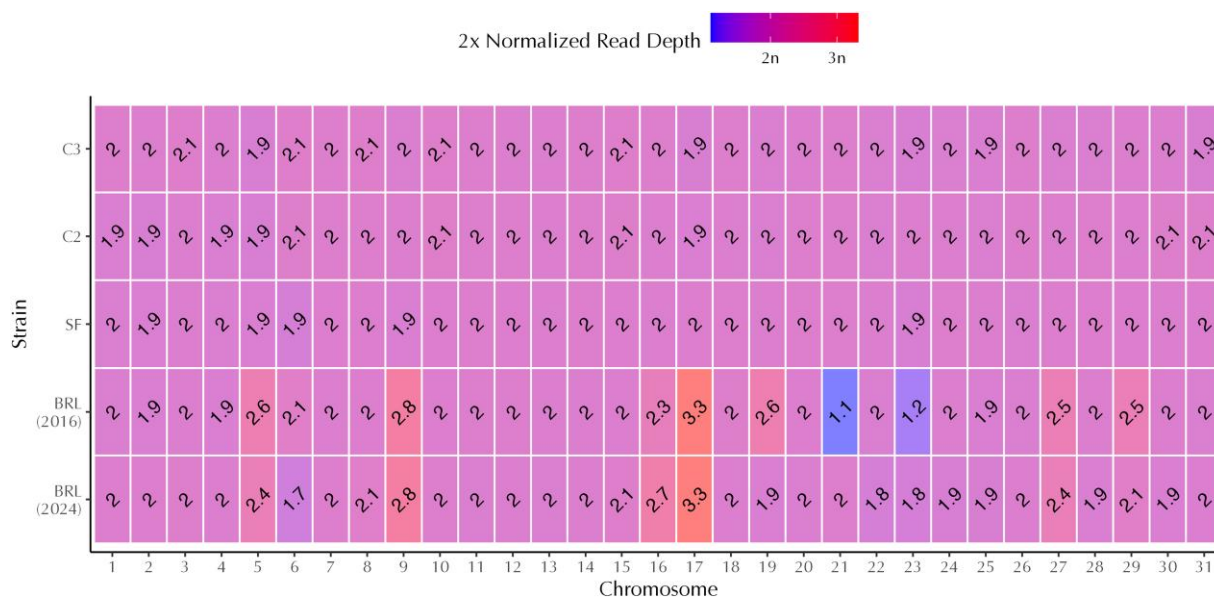
**Fig. 4.** Circular plot of *Lotmaria passim* BRL (2024) genome assembly. Radial segments correspond to the 31 nuclear chromosomes and are numbered by size from largest to smallest. Sequencing depth is presented within each segment. The gray points show base-level read depths, subsampled at 100-bp intervals, while the blue line trace represents a 1 Kb moving average, and the red circles represent chromosome-level medians. Concentric yellow, orange, and red dashed lines represent 50%, 100%, and 150% of the median coverage depth, which corresponds to the expected depths for monosomic, disomic, and trisomic chromosomes, respectively. The pink link between chromosome 5 and chromosome 6 indicates high similarity (98.62%) between these two regions, suggesting a chromosomal duplication event previously occurred. The faint scatter of gray points roughly half the read depth of the chromosomes overall (referred to as the 'secondary trace' in the text) suggests that heterozygous sites are conspicuously absent from chromosome 21, suggesting that our strain harbors two identical haplotypes of this chromosome.

(2024) has again shifted to a mixture of diploids with two distinct and two identical copies of the chromosome.

The last and arguably most remarkable evidence of *Lotmaria passim*'s plasticity in some lies within chromosome 21. During our assembly and further analysis of the BRL (2024) reference genome, we noticed an apparent lack of the base-level read depths (secondary trace) hovering at roughly half the read depth of the chromosomes overall which are apparent in the other 30 chromosomes (gray points; Fig. 4, Supplementary Fig. 1). This suggests that the several heterozygous sites in each of the other 30 chromosomes in the genome are conspicuously absent in chromosome 21, leading us to believe that a single monosomic copy was recently duplicated to bring chromosome 21 to a state of disomy. We found that chromosome 21 in the BRL (2016) strain appeared to be monosomic as reflected by its 1.1 $\times$  normalized read depth, while the BRL (2024) assembly appears disomic (Fig. 5). We hypothesize that since the BRL (2016) assembly, the BRL isolate underwent a duplication of the remaining copy of chromosome 21 during axenic cell culture expansion, which thereby restored chromosome 21 to disomy in subsequent cell cultures that were

eventually used for sequencing of the BRL (2024) strain. The uniquely low heterozygosity observed in chromosome 21 of the BRL (2024) assembly would thus be explained by such a recent duplication event and by the limited time these cultures are maintained in an active state over which unique mutations could accrue; cultures are frozen as stocks and grown only when expansions are needed. Both the 2 $\times$  normalized read depth (Fig. 5) and lack of variants (Supplementary Fig. 2) support this hypothesis, although karyotyping will likely be necessary for unambiguous inference.

One potential explanation for these shifts within the earlier BRL (2016) and current BRL (2024) populations could be due to the intervening period during which the line has been maintained in culture, which includes approximately six months (~500 generations) and several cryopreservation events in our laboratory, as well as shifts from antibiotic-supplemented Schneider's medium (Schwarz et al. 2015) used during initial cultivation to modified FPF medium (Salathé et al. 2012) which we currently use to support greater cell densities. It is widely accepted that relatively rapid changes in chromosome copy number are considered a



**Fig. 5.** Heatmap of normalized sequencing depth for five *Lotmaria passim* genome assemblies. The x-axis denotes the nuclear chromosome number and the y-axis denotes the *Lotmaria passim* strain of origin for each assembly. Shading and text within each box depict the median sequencing depth of each chromosome when mapped to the *Lotmaria passim* BRL (2024) assembly after normalization to the median chromosome-level depth of the corresponding assembly. These depths are used as a tentative indicator of somy (i.e. chromosome copy number), with a 2x normalized read depth of 1x, 2x, and 3x denoting a monosomic, disomic, or trisomic chromosome, respectively. We display doubled normalized depth values such that the number on the plot corresponds to chromosome copy number. *Lotmaria passim* BRL (2016) normalized read depth indicates the loss of one haploid chromosome in comparison to strains C2, C3, and SF, while *Lotmaria passim* BRL (2024) normalized read depth provides evidence for the recent duplication of chromosome 21, returning it to a state of disomy in this assembly.

mechanism by which trypanosomatids adapt to changing environments (Sterkers et al. 2012; Reis-Cunha et al. 2018; Albanaz et al. 2023). This adaptation is evidenced by evolution experiments in *Leishmania*, suggesting that both time and changes in growth conditions may have contributed to the inferred genomic changes (Ponte-Sucre et al. 2017). Microevolution assays performed on *Trypanosoma cruzi* also evidence these adaptations, suggesting that continuous time in culture led to ‘genome erosion’ events as tetraploid hybrids shifted toward stable populations of triploid hybrids (Matos et al. 2022).

### Variation in somy across strains

To investigate patterns of aneuploidy in *Lotmaria passim* outside of the BRL strain, we conducted parallel analyses on three additional, independently isolated lines: a 2016 sequencing of the SF strain used for the original *Lotmaria passim* draft assembly (Runckel et al. 2014) and the recently sequenced Spanish isolates named ‘C2’ and ‘C3’ (Buendía-Abad et al. 2021). In contrast to the eccentricities in copy number suggested by the two assemblies of the BRL strain (2016 sequencing effort (GenBank Accession ID: GCA\_002216525.1) and 2024 sequencing effort (GenBank Accession ID: GCA\_037349495.1)), none of the three other strains sequenced showed convincing evidence of deviation from expectations for disomy at any chromosome. Although allele frequencies exhibited some pattern of tetrasomy for chromosomes 5 and 6, analysis sequencing depth suggested disomy (Supplementary Fig. 1); the anomalous allele frequencies likely reflect difficulty distinguishing these two highly similar chromosomes using short reads. These findings leave the extent of aneuploidy in natural populations and its ramifications for parasite fitness unclear. Given that the freshly isolated, low passage ‘C1’ strain resulted in greater host mortality as opposed to the more extensively cultured strain SF in experimentally inoculated honey bees (Buendía-Abad et al. 2021), understanding how

both laboratory cultivation and other environmental shifts affect the genome evolution and within-host dynamics of this parasite is warranted and has potential to elucidate mechanisms of adaptation in this host-parasite system.

Moreover, the extent of aneuploidy exhibited in the BRL strain lies well within ranges reflected by read-depth analyses of *Trypanosoma cruzi* discrete typing units TcII and TcIII (Reis-Cunha et al. 2018) and found in over 20 other trypanosomatids of varying ages post-isolation (Albanaz et al. 2023). Paired with the evidence for a past duplication of chromosomes 5 and 6, it is likely that copy numbers are flexible in naturally circulating *Lotmaria passim* populations as well. In addition, within-species differences in ploidy exist across isolates of the closely related *Leptomonas pyrrocoris* (Flegontov et al. 2016). Sequencing of additional *Lotmaria passim* strains would be necessary to gain a more complete representation of how somy varies within and across populations, as well as its environmental correlates and consequences for parasite infectivity and reproduction.

### Conclusions

The *Lotmaria passim* BRL (2024) genome is one of the most comprehensive genomic resources available for monoxenous trypanosomatids within the subfamily Leishmaniinae and greatly improves upon the previously available reference assembly (Runckel et al. 2014), providing a high-quality resource for future work that promises insights into the evolutionary changes leading to specialization within honey bees. Our assembly provides additional examples of aneusomies within trypanosomatids and suggests further study of the causes and consequences of variation in somy in monoxenous species is necessary. The honey bee-*Lotmaria passim* system offers many parallels to the challenges faced by other Leishmaniinae during infection of mammals and exposure to

antiparasitic drugs and will be an important tool for understanding mechanisms of infection and adaptation in their medically important dioxenous relatives.

## Data availability

All *Lotmaria passim* BRL (2024) Hi-C and HiFi raw reads are available on NCBI GenBank (accession ID: SRX22798691 and SRX22798690) and the assembled chromosomes (accession ID: GCA\_037349495.1), maxicircle, and minicircles are listed under BioProject PRJNA1049372. The annotated genome will be available on NCBI GenBank under BioProject PRJNA1049372 and on FigShare under the filename “*Lotmaria.passim\_annotation.gff3*.” Detailed methods regarding both the assembly and annotation can be found herein within the manuscript. Supplementary figures can be found on FigShare at <https://doi.org/10.25387/g3.27121227>.

## Acknowledgments

The authors would like to thank Amanda Albanaz for proteome annotations of *Crithidia mellificae*, *Crithidia bombi*, and *Crithidia expoeki* and the anonymous reviewers for their comprehensive and constructive feedback which was used to improve the manuscript. Use of specific materials and equipment is for documentation purposes only and does not imply endorsement of these products.

## Funding

This project was supported by U.S. Department of Agriculture National Institute of Food and Agriculture Grant 2020-67013-31861 to JDE and ECPY, U.S. Department of Agriculture National Institute of Food and Agriculture Postdoctoral Fellowship 2022-67012-37482 to ECPY, an Eva Crane Trust Grant to JDE and ECPY, the U.S. Department of Agriculture Agricultural Research Service Beltsville Bee Research Laboratory in-house funds, and U.S. National Science Foundation Grant DEB-2225083 to CM. In addition, this material is based upon work supported by the U.S. National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-2236417 to LMM. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Funding agencies had no role in the associated experimental design, data collection, interpretation, or publication of these results.

## Conflicts of interest

The authors declare no conflicts of interest.

## Literature cited

Akdemir KC, Chin L. 2015. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* 16(1):198. doi:10.1186/s13059-015-0767-1.

Albanaz ATS, Carrington M, Frolov AO, Ganyukova AI, Gerasimov ES, Kostygov AY, Lukeš J, Malysheva MN, Votýpka J, Zakharova A, et al. 2023. Shining the spotlight on the neglected: new high-quality genome assemblies as a gateway to understanding the evolution of trypanosomatidae. *BMC Genomics.* 24(1):471. doi:10.1186/s12864-023-09591-z.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.

Andrews S, Kreuger F, Segonds-Pichon A, Biggins L, Kreuger C, Wingett S. 2012. FastQC A Quality Control tool for High Throughput Sequence Data. [accessed 2024 May 30]. Available from <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.

Arismendi N, Bruna A, Zapata N, Vargas M. 2016. PCR-specific detection of recently described *Lotmaria passim* (Trypanosomatidae) in Chilean apiaries. *J Invertebr Pathol.* 134:1–5. doi:10.1016/j.jip.2015.12.008.

Bao W, Kojima KK, Kohany O. 2015. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 6(1):11. doi:10.1186/s13100-015-0041-9.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27(2):573–580. doi:10.1093/nar/27.2.573.

Buendía-Abad M, García-Palencia P, de Pablos LM, Alunda JM, Osuna A, Martín-Hernández R, Higes M. 2022. First description of *Lotmaria passim* and *Crithidia mellificae* haptomonad stages in the honeybee hindgut. *Int J Parasitol.* 52(1):65–75. doi:10.1016/j.ijpara.2021.06.005.

Buendía-Abad M, Higes M, Martín-Hernández R, Barrios L, Meana A, Fernández A, Osuna A, De Pablos LM. 2021. Workflow of *Lotmaria passim* isolation: experimental infection with a low-passage strain causes higher honeybee mortality rates than the PRA-403 reference strain. *Int J Parasitol Parasites Wildl.* 14:68–74. doi:10.1016/j.ijppaw.2020.12.003.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics.* 10(1):421. doi:10.1186/1471-2105-10-421.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. Trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–1973. doi:10.1093/bioinformatics/btp348.

Castelli L, Branchiccela B, Invernizzi C, Tomasco I, Basualdo M, Rodríguez M, Zunino P, Antúnez K. 2019. Detection of *Lotmaria passim* in Africanized and European honey bees from Uruguay, Argentina and Chile. *J Invertebr Pathol.* 160:95–97. doi:10.1016/j.jip.2018.11.004.

Chen N. 2004. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* 5(1):4.10.1–4.10.14. doi:10.1002/0471250953.bi0410s05.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta.* 2(2):e107. doi:10.1002/imt2.107.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Cornman RS, Tarpy DR, Chen Y, Jeffreys L, Lopez D, Pettis JS, vanEngelsdorp D, Evans JD. 2012. Pathogen webs in collapsing honey bee colonies. *PLoS One.* 7(8):e43562. doi:10.1371/journal.pone.0043562.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience.* 10(2):giab008. doi:10.1093/gigascience/giab008.

de Paula JC, Olmedo PG, Gómez-Moracho T, Buendía-Abad M, Higes M, Martín-Hernández R, Osuna A, de Pablos LM. 2024. Promastigote EPS secretion and haptomonad biofilm formation as evolutionary adaptations of trypanosomatid parasites for colonizing honeybee hosts. *NPJ Biofilms Microbiomes.* 10(1):27. doi:10.1038/s41522-024-00492-x.

Dainat J, Hereñú D, Pucholt P. 2020. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF. doi:10.5281/zenodo.11106497. <https://zenodo.org/records/11106497>.

- Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for Hi-C. Contact Maps with Unlimited Zoom. *cells*. 3(1):99–101. doi:10.1016/j.cels.2015.07.012.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7(10):e1002195. doi:10.1371/journal.pcbi.1002195.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):238. doi:10.1186/s13059-019-1832-y.
- Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, Filatov D, Flegontova O, Gerasimov ES, Hlaváčková J, et al. 2016. Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep*. 6(1):23704. doi:10.1038/srep23704.
- Gerasimov ES, Gasparyan AA, Afonin DA, Zimmer SL, Kraeva N, Lukeš J, Yurchenko V, Kolesnikov A. 2021. Complete minicircle genome of *Leptomonas pyrrhocoris* reveals sources of its non-canonical mitochondrial RNA editing events. *Nucleic Acids Res*. 49(6):3354–3370. doi:10.1093/nar/gkab114.
- Gómez-Moracho T, Buendía-Abad M, Benito M, García-Palencia P, Barrios L, Bartolomé C, Maside X, Meana A, Jiménez-Antón MD, Ollás-Molero AI, et al. 2020. Experimental evidence of harmful effects of *Crithidia mellificae* and *Lotmaria passim* on honey bees. *Int J Parasitol*. 50(13):1117–1124. doi:10.1016/j.ijpara.2020.06.009.
- Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E. 2013. TIGRFAMs and genome properties in 2013. *Nucleic Acids Res*. 41(D1):D387–D395. doi:10.1093/nar/gks1234.
- Hoff KJ, Stanke M. 2019. Predicting genes in single genomes with AUGUSTUS. *Curr Protoc Bioinformatics*. 65(1):e57. doi:10.1002/cpbi.57.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 12(1):491. doi:10.1186/1471-2105-12-491.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780. doi:10.1093/molbev/mst010.
- Knaus BJ, Grünwald NJ. 2017. Vcfr: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resour*. 17(1):44–53. doi:10.1111/1755-0998.12549.
- Knaus BJ, Grünwald NJ. 2018. Inferring variation in copy number using high throughput sequencing data in R. *Front Genet*. 9:123. doi:10.3389/fgene.2018.00123.
- Koch H, Brown MJF, Stevenson PC. 2017. The role of disease in bee foraging ecology. *Curr Opin Insect Sci*. 21:60–67. doi:10.1016/j.cois.2017.05.008.
- Kostygov AY, Albanaz ATS, Butenko A, Gerasimov ES, Lukeš J, Yurchenko V. 2024. Phylogenetic framework to explore trait evolution in trypanosomatidae. *Trends Parasitol*. 40(2):96–99. doi:10.1016/j.pt.2023.11.009.
- Krivtseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. *Nucleic Acids Res*. 47(D1):D807–D811. doi:10.1093/nar/gky1053.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359. doi:10.1038/nmeth.1923.
- Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 30(22):3276–3278. doi:10.1093/bioinformatics/btu531.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol*. 21(6):1095–1109. doi:10.1093/molbev/msh112.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol*. 62(4):611–615. doi:10.1093/sysbio/syt022.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. doi:10.48550/arXiv.1303.3997.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34(18):3094–3100. doi:10.1093/bioinformatics/bty191.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.
- Liu Q, Lei J, Darby AC, Kadowaki T. 2020. Trypanosomatid parasite dynamically changes the transcriptome during infection and modifies honey bee physiology. *Commun Biol*. 3(1):51. doi:10.1038/s42003-020-0775-x.
- Lukeš J, Butenko A, Hashimi H, Maslov DA, Votýpka J, Yurchenko V. 2018. Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. *Trends Parasitol*. 34(6):466–480. doi:10.1016/j.pt.2018.03.002.
- MacInnis CI, Luong LT, Pernal SF. 2023. A tale of two parasites: responses of honey bees infected with *Nosema ceranae* and *Lotmaria passim*. *Sci Rep*. 13(1). doi:10.1038/s41598-023-49189-9.
- Maslov DA, Opperdoes FR, Kostygov AY, Hashimi H, Lukeš J, Yurchenko V. 2019. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology*. 146(1):1–27. doi:10.1017/S0031182018000951.
- Matos GM, Lewis MD, Talavera-López C, Yeo M, Grisard EC, Messenger LA, Miles MA, Andersson B. 2022. Microevolution of *Trypanosoma cruzi* reveals hybridization and clonal mechanisms driving rapid genome diversification. *eLife*. 11:e75237. doi:10.7554/eLife.75237.
- McGhee RB, Cosgrove WB. 1980. Biology and physiology of the lower Trypanosomatidae. *Microbiol Rev*. 44(1):140–173. doi:10.1128/mr.44.1.140-173.1980.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20(9):1297–1303. doi:10.1101/gr.107524.110.
- Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, et al. 2020. Pfam: the protein families database in 2021. *Nucleic Acids Res*. 49(D1):D412–D419. doi:10.1093/nar/gkaa913.
- Morimoto T, Kojima Y, Yoshiyama M, Kimura K, Yang B, Peng G, Kadowaki T. 2013. Molecular detection of protozoan parasites infecting *Apis mellifera* colonies in Japan. *Environ Microbiol Rep*. 5(1):74–77. doi:10.1111/j.1758-2229.2012.00385.x.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 29(22):2933–2935. doi:10.1093/bioinformatics/btt509.
- Pagès H, Aboyoum P, Gentleman R, DebRoy S. 2024. Biostrings: Efficient manipulation of biological strings. [accessed 2024 May 30]. <http://bioconductor.org/packages/Biostrings/>.
- Paradis E, Schliep K. 2019. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*. 35(3):526–528. doi:10.1093/bioinformatics/bty633.

- Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Res.* 9:ISCB Comm J-304. doi:10.12688/f1000research.23297.2.
- Ponte-Sucre A, Gamarro F, Dujardin J-C, Barrett MP, López-Vélez R, García-Hernández R, Pountain AW, Mwenechanya R, Papadopoulos B. 2017. Drug resistance and treatment failure in leishmaniasis: a 21st century challenge. *PLoS Negl Trop Dis.* 11(12):e0006052. doi:10.1371/journal.pntd.0006052.
- Ravoet J, Maharramov J, Meeus I, De Smet L, Wenseleers T, Smagghe G, de Graaf DC, Li Y. 2013. Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellificae* as a new contributory factor to winter mortality. *PLoS ONE.* 8(8):e72443. doi:10.1371/journal.pone.0072443.
- Ravoet J, Schwarz RS, Descamps T, Yañez O, Tozkar CO, Martín-Hernández R, Bartolomé C, De Smet L, Higes M, Wenseleers T. 2015. Differential diagnosis of the honey bee trypanosomatids *Crithidia mellificae* and *Lotmaria passim*. *J Invertebr Pathol.* 130:21–27. doi:10.1016/j.jip.2015.06.007.
- R Development Core Team. 2009. R: A language and environment for statistical computing. <http://www.R-project.org>.
- Reis-Cunha JL, Valdivia HO, Bartholomeu DC. 2018. Gene and chromosomal copy number variations as an adaptive mechanism towards a parasitic lifestyle in Trypanosomatids. *Curr Genomics.* 19(2): 87–97. doi:10.2174/1389202918666170911161311.
- Robinson JT, Thorvaldsdottir H, Turner D, Mesirov JP. 2023. Igv.js: an embeddable JavaScript implementation of the integrative genomics viewer (IGV). *Bioinformatics.* 39(1):btac830. doi:10.1093/bioinformatics/btac830.
- Runckel C, DeRisi J, Flenniken ML. 2014. A draft genome of the honey bee trypanosomatid parasite *Crithidia mellificae*. *PLoS One.* 9(4): e95057. doi:10.1371/journal.pone.0095057.
- Salathé R, Tognazzo M, Schmid-Hempel R, Schmid-Hempel P. 2012. Probing mixed-genotype infections I: extraction and cloning of infections from hosts of the Trypanosomatid *Crithidia bombi*. *PLoS One.* 7(11):e49046. doi:10.1371/journal.pone.0049046.
- Schmid-Hempel P. 2001. On the evolutionary ecology of host-parasite interactions: addressing the question with regard to bumblebees and their parasites. *Naturwissenschaften.* 88(4):147–158. doi:10.1007/s001140100222.
- Schmid-Hempel P, Aebi M, Barribeau S, Kitajima T, Plessis L du, Schmid-Hempel R, Zoller S. 2018. The genomes of *Crithidia bombi* and *C. expoeki*, common parasites of bumblebees. *PLoS One.* 13(1): e0189738. doi:10.1371/journal.pone.0189738.
- Schwarz RS, Baughan GR, Murphy CA, Ravoet J, de Graaf DC, Evans JD. 2015. Characterization of two species of trypanosomatidae from the honey bee *Apis mellifera*: *Crithidia mellificae* Langridge and McGhee, and *Lotmaria passim* n. Gen., n. sp. *J Eukaryot Microbiol.* 62(5):567–583. doi:10.1111/jeu.12209.
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. *Gene Prediction: Methods and Protocols.* New York, NY: Springer. p. 227–245.
- Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* 16(1):259. doi:10.1186/s13059-015-0831-x.
- Shanmugasundram A, Starns D, Böhme U, Amos B, Wilkinson PA, Harb OS, Warrenfeltz S, Kissinger JC, McDowell MA, Roos DS, et al. 2023. TriTrypDB: an integrated functional genomics resource for kinetoplastida. *PLoS Negl Trop Dis.* 17(1):e0011058. doi:10.1371/journal.pntd.0011058.
- Sim SB, Corpuz RL, Simmonds TJ, Geib SM. 2022. HiFiAdapterFilter, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. *BMC Genomics.* 23(1):157. doi:10.1186/s12864-022-08375-1.
- Steinbiss S, Silva-Franco F, Brunk B, Foth B, Hertz-Fowler C, Berriman M, Otto TD. 2016. Companion: a web server for annotation and analysis of parasite genomes. *Nucleic Acids Res.* 44(W1): W29–W34. doi:10.1093/nar/gkw292.
- Sterkers Y, Lachaud L, Bourgeois N, Crobu L, Bastien P, Pagès M. 2012. Novel insights into genome plasticity in Eukaryotes: mosaic aneuploidy in *Leishmania*. *Mol Microbiol.* 86(1):15–23. doi:10.1111/j.1365-2958.2012.08185.x.
- Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. 2021. The Dfam community resource of transposable element families, sequence models, and genome annotations. *Mob DNA.* 12(1):2. doi:10.1186/s13100-020-00230-y.
- Syberg-Olsen MJ, Garber AI, Keeling PJ, McCutcheon JP, Husnik F. 2022. Pseudofinder: detection of pseudogenes in prokaryotic genomes. *Mol Biol Evol.* 39(7):msac153. doi:10.1093/molbev/msac153.
- Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 28(5):511–515. doi:10.1038/nbt.1621.
- UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* 43(D1):D204–D212. doi:10.1093/nar/gku989.
- Vejnovic B, Stevanovic J, Schwarz RS, Aleksic N, Mirilovic M, Jovanovic NM, Stanimirovic Z. 2018. Quantitative PCR assessment of *Lotmaria passim* in *Apis mellifera* colonies co-infected naturally with *Nosema ceranae*. *J Invertebr Pathol.* 151:76–81. doi:10.1016/j.jip.2017.11.003.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 9(11):e112963. doi:10.1371/journal.pone.0112963.
- Wickham H, Navarro D, Pederson TL. 2016. *ggplot2: Elegant Graphics for Data Analysis.* 3rd ed. Springer-Verlag New York.
- Xu G, Palmer-Young E, Skyrms K, Daly T, Sylvia M, Averill A, Rich S. 2018. Triplex real-time PCR for detection of *Crithidia mellificae* and *Lotmaria passim* in honey bees. *Parasitol Res.* 117(2):623–628. doi:10.1007/s00436-017-5733-2.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 8(1): 28–36. doi:10.1111/2041-210X.12628.
- Yuan X, Sun J, Kadowaki T. 2024. Aspartyl protease in the secretome of honey bee trypanosomatid parasite contributes to infection of bees. *Parasit Vectors.* 17(1):60. doi:10.1186/s13071-024-06126-7.
- Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 39(1):btac808. doi:10.1093/bioinformatics/btac808.