

RESEARCH

Open Access



Chromosome-level genome assembly of trypanosomatid parasite *Lotmaria passim* links chromosome duplication and divergence with infection of honey bees

Anthony Nearman^{1†} , Anzhelika Butenko^{2,3,4}, Jay D Evans^{1*} and Evan C Palmer-Young^{1*†}

Abstract

Background The protist family Trypanosomatidae includes parasites of insects, vertebrates, plants, and even other unicellular eukaryotes. The genomes of these species harbor clues to the evolution of parasitism, adaptation to new hosts, and infection of mammals. We present an analysis of a chromosome-level genome assembly of *Lotmaria passim*, the most prevalent known trypanosomatid of honey bees, linking genome sequence and organization to gene expression and infection of bees.

Results The genome showed a high degree of synteny with assemblies of other trypanosomatids and especially to the closely related *Leptomonas pyrrhocoris*. It included four copies of chromosomes that shared ancestry with the tetrasomic *Leishmania* Chromosome 31 and are consistently supernumerary throughout Trypanosomatidae. However, these chromosomes showed lower similarity to *L. passim* relatives than did the genome overall, with sufficient variation across haplotypes to distinguish two separate disomic chromosomes. Transcriptomic analyses showed that these chromosomes are enriched in genes upregulated during bee infection, and each include five paralogs of the GP63 gene implicated in infection of both insects and mammals. Patterns of expression in bees suggested decreased protein synthesis, a shift from carbohydrate- to amino acid-based metabolism, and reduced cell motility in bee guts versus cell culture. In contrast, genes involved in cell adhesion were upregulated, consistent with the importance of attachment to insect tissue in this species and the family overall.

Conclusions Our analysis links differentiation of a conserved supernumerary chromosome with infection of bees, parallel to this chromosome's role in *Leishmania* infection of mammals and linking chromosome-level changes with adaptation to new hosts.

[†]Anthony Nearman and Evan C Palmer-Young contributed equally.

*Correspondence:

Jay D Evans

jay.evans@usda.gov

Evan C Palmer-Young

evan.palmeryoung@gmail.com

Full list of author information is available at the end of the article



This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Keywords Aneuploidy, Trypanosomatid ancestral supernumerary chromosome, Gene duplication, Phylogenomics, Polycistronic genome organization, Post-transcriptional gene regulation, Differential gene expression, Host-parasite interactions, *Crithidia*

Introduction

The trypanosomatids are an intriguing and medically important family of protists that exploit a variety of ecological niches. They include obligate monoxenous (single-host) parasites of insects and dixenous (two-host) insect-vectored parasites of plants and vertebrates [1–3], namely the *Leishmania* and *Trypanosoma* species that cause considerable neglected diseases in tropical and subtropical regions [4–6]. The factors that enable the radiation of trypanosomatids into novel host environments are therefore of basic and applied interest [2, 7].

The sequencing of genomes from trypanosomatids across this spectrum of lifestyles has begun to reveal the phylogenetic relationships between species, changes associated with the transition to parasitism, and adaptations to different niches [2, 8–10]. Several genetic features have emerged from the species investigated thus far. These include a generally conserved genome structure with high degree of synteny among distantly related species, the arrangement and transcription of genes in long polycistronic blocks, and predominantly post-transcriptional regulation of gene expression [11, 12].

Many species also exhibit duplication of chromosome segments and variation in chromosome copy number, or aneusomy [9, 13, 14]. Although such duplications create a potentially costly excess of the genes present in the supernumerary region, they can also benefit fitness through increased dosage of specific advantageous genes [14]. This can facilitate adaptation to ecological and evolutionary pressures, such as exposure to toxins and new host environments [14–17]. One remarkable example is the *Leishmania* Chromosome 31, which is consistently tetrasomic (i.e., present as four copies) across species and strains in this genus [14, 17]. It is also enriched in genes that are upregulated during infection of vertebrates—including those encoding amastins as well as amino acid transporters—suggesting that this duplication aided the ability to colonize these new hosts [16]. A recent investigation of over 800 trypanosomatid genome assemblies revealed that scaffolds and chromosomes syntenic with Chromosome 31 are in fact tetrasomic across a wide range of mono- and dixenous species, suggesting that this aneusomy is ancestral in the Trypanosomatidae family rather than a specific adaptation of *Leishmania* [14]. However, its importance for the infection of other hosts and insect vectors remains unclear.

Among the monoxenous trypanosomatids, several species are associated with bees [18]. *Crithidia bombi* and *C. expoeki*, the main species identified in bumble bees, have

emerged as models of host-parasite ecology and evolution and revealed the context- and life stage-dependent effects of trypanosomatid infection on insects [19–21]. *Lotmaria passim*, on the other hand, is the most prevalent trypanosomatid species known in honey bees [18]. The parasite colonizes the walls of the honey bee ileum and rectum [18, 22], with fecal-oral transmission inferred from its presence in honey bee feces [23]. It exhibits a global distribution, high prevalence, negative effects on bee survival, and an association with collapsing colonies [24–33], suggesting its importance as a threat to bee health. In addition, *L. passim*'s bee gut niche offers several parallels with the intracellular niche of *Leishmania* in vertebrates [2, 34]. These include exposure within bees to elevated temperatures similar to those of the mammalian bloodstream [35, 36]; and establishment in the acidic environment of the bee hindgut [37], where pH resembles that in the acidic phagolysosome colonized by *Leishmania* amastigotes [38, 39]. Hence, its adaptations to this niche could provide insights into trypanosomatids' infection of mammals.

Our group recently published a high-quality, chromosome-level assembly of the BRL (Bee Research Lab) type strain (American Type Culture Collection isolate PRA-422, GCA_037349495.1) of *L. passim* [40]. The culture was isolated from the ileum of adult female *Apis mellifera* at the United States Department of Agriculture Bee Research Lab in Beltsville, Maryland, USA in September 2012 [18]. Compared to the draft genome of strain SF (San Francisco, ATCC isolate PRA-403) published ten years prior—before the differentiation of *L. passim* from its relative *C. mellificae* [41]—the new assembly is more contiguous (number of nuclear genome scaffolds decreased from 2801 to 31), complete (the representation of Euglenozoa BUSCO genes increased from 90 to 98.5%), contains mitochondrial genome sequences, and enabled discernment of chromosome-level aneusomy [40]. To our knowledge, it represents the first chromosome-level assembly for a species in the Crithidiatae clade comprised of *Lotmaria*, *Leptomonas*, and *Crithidia* [8]; the first for a monoxenous species in the Leishmaniinae subfamily; and the second for a monoxenous trypanosomatid (after *Angomonas deanei* [42]). This yielded a more complete understanding of the genome's structure, while providing a framework to investigate the relationship between gene arrangement, expression, and function. One of the striking elements of the assembly was the presence of a paralogous chromosome pair represented by

Chromosomes 5 and 6, the origins and significance of which remained unclear [40].

To discover the distinguishing features of the *L. passim* genome—including its paralogous chromosome pair—and their relevance to infection of honey bees, we compared the structure of the genome with that of related trypanosomatids, quantified spatial patterns of gene expression within and across chromosomes, and analyzed functional changes in gene expression of parasites in bee guts relative to cell cultures. Our results highlight the evolution of *L. passim*'s two paralogous chromosomes and their orthology with *Leishmania* spp. supernumerary Chromosome 31. We also link genes on these chromosomes with infection of bees and point to the importance of attachment to host tissue in the parasite's life cycle.

Methods

Syntenic analysis

The genome of the *L. passim* BRL type strain (maintained in the American Type Culture Collection as isolate "PRA-422", GenBank assembly GCA_037349495.1) was previously sequenced using a combination of Pac-Bio HiFi and Illumina Hi-C technologies, assembled to 31 nuclear chromosomes, and annotated with 10,288 genes [40].

Conservation of gene order (synteny) between *L. passim* and other Leishmaniinae species with high-quality assemblies (*Leptomonas seymouri*; *Crithidia bombi*, *C. expoeki*, and *C. fasciculata*; *Leishmania major* and *L. braziliensis* (Supplementary Table S1)), including Crithidia genomes with <250 scaffolds and one chromosome-level assembly from each subgenus of *Leishmania*, was quantified with SyMAP v5.9.9 [43] using default parameters. SyMAP raw data was exported and visualized using the R package circlize v0.4.16 [44].

To examine telomeric sequences, tandem repeats found at chromosome termini were identified and counted using the Telomere Identification toolKit [45]. We used the 'explore' function to identify repeated sequences from 5 to 15 bp long; and the 'search' function to detect matches to the canonical sequence "CCCTAA", used as a marker for trypanosomatid telomeres [46], within 1% of each chromosome's length from either end.

Phylogenetic reconstruction

To define the phylogenetic relationships between *L. passim* and other trypanosomatids, we constructed a tree based on predicted proteins from 15 *Lotmaria*, *Leptomonas*, *Crithidia*, and *Leishmania* species in the Leishmaniinae subfamily (Supplementary Table S2). For species without publicly available annotations, we used those made in a previous phylogenomic analysis [8]. Proteins were aligned using MAFFT through Orthofinder v3.0.1b1, resulting in 2,530 single-copy orthologs from which a phylogenetic tree was inferred using fasttree [47].

Orthofinder was run with default parameters except -M msa to enable multisequence alignment and tree inference. Tree visualization was performed using R packages ape v5.8.1 and ggtree v3.16.3 [48, 49].

Spatial analysis of genes and gene expression

To quantify the spatial clustering of genes by strand within each chromosome, we created a matrix for all gene pairs and their distances apart along the chromosome. We used this to model the correlation between $\log_{10}(\text{pairwise distance})$ (included as both a linear and quadratic term) and the probability of genes being on the same strand using a binomial model with chromosome as a random effect. This and subsequent models were implemented in R version 4.4 using package glmmTMB v1.1.12 to construct models, car v3.1.3 to evaluate significance of predictor terms, emmeans v1.11.2.8 to compute marginal means for different values of predictor variables, and ggplot2 v3.5.2 to visualize results [50–53]. We also used a binomial model to compare the proportion of features annotated as pseudogenes on the paralogous Chromosomes 5 and 6 vs. the rest of the genome.

We then combined our new genome annotation with publicly available transcriptomic data to examine chromosome-scale patterns of gene expression and up- and downregulation in honey bee hindguts relative to parasite cell cultures. The transcriptomic data (originally described in [32]) consisted of 30 samples from the type strain 'SF' (ATCC isolate 'PRA-403' [41]) grown in antibiotic-supplemented 'FPFB' medium [54] and gut samples of parasite-inoculated honey bees (*Apis mellifera*) at 7, 12, 20, and 27 d post-inoculation ($n=6$ replicate libraries per group). Reads from the associated SRA files were mapped to the *Lotmaria passim* genome (GCA_037349495.1) using the short read alignment software STAR v2.7.11b [55] using default alignment parameters for unpaired reads and summarized by gene using featureCounts v2.0.6 [56]. Reads that mapped to multiple genes were fractionally distributed among each potential gene from which they could have originated (featureCounts -M --fraction). One of the replicates for the 7 d time point was excluded due to low read counts. Differential expression was analyzed using R package DESeq2 v1.48.2 [57] to compare expression in the bee gut relative to cell culture at each time point post-inoculation (see Supplementary Methods for additional details). Differences in proportions of differentially expressed genes between the paralogous chromosomes and the rest of the nuclear genome were evaluated using binomial models for five different differential expression categories: Statistically significant upregulation, downregulation, or differential expression in either direction with respect to cultured cells (adjusted p -value < 0.05, any fold change); and strong (adjusted p -value < 0.05 and $|\text{fold-change}| > 2$)

up- and downregulation. Because the gene expression patterns within each chromosome were generally consistent across the four time points, we fit a single model for each expression category with chromosome type (Chromosomes 5 and 6 vs. rest of genome) as a fixed predictor and time point as a random effect. We also compared the absolute expression level (measured as $\log_2$ (transcripts per million (TPM)), pooled across all libraries) between genes on Chromosomes 5 and 6 vs. other chromosomes.

The amount of variance in expression (i.e., gene length-corrected read count) explained by same-stranded gene clusters of more than 10 consecutive genes was tested using linear mixed models. We fit separate models for each of the five treatment groups, then extracted the proportion of variance explained by the gene cluster random effect term (see Supplementary Methods). We also analyzed correlations in absolute and differential expression between adjacent genes on the same vs. opposite strand of the chromosome (see Supplementary Methods).

Functional gene annotation and patterns of expression

We conducted functional analyses of differential gene expression in the bee gut vs. cell culture using Gene Set Enrichment Analysis (GSEA [58]) of Gene Ontology (GO [59, 60]) and Kyoto Encyclopedia of Genes and Genomes (KEGG [61]) terms. Genes were assigned tentative GO terms by submitting the annotated protein sequences to the PANNZER2 server [62]. Terms were filtered to retain only predictions with a positive prediction value of >0.5 , resulting in annotations for 2,457 of the original 10,288 submitted sequences. Parent terms in the GO hierarchy implied by each prediction were added using the build-GOmap function of R package ClusterProfiler v3.21 [63]. The resulting set of terms was then filtered using FunTaxIS-lite to remove terms unlikely to occur in trypanosomatids [64]. Annotated genes were similarly assigned KEGG functions based on protein sequences and annotated to associated pathways using BlastKOALA [65], which inferred KEGG terms for 2,693 of the submitted protein sequences. The resulting list of terms was filtered to retain only pathways annotated in reference kinetoplastid genomes present in the KEGG database.

Gene set enrichment analysis, which uses a rank-based enrichment score to identify gene sets whose genes are overrepresented at the top and bottom of a ranked gene list [58], was implemented on gene sets associated with each GO term or KEGG pathway using ClusterProfiler [63] on all terms associated with at least 15 genes. Genes were ranked within each time point using the test statistic from the differential expression estimate for bee gut-derived relative to cultured cells (i.e., the ratio of the estimated $\log_2$ -fold change to the standard error of the estimate) as the signal-to-noise ratio (details in

Supplementary Methods). We discuss only gene sets with adjusted p-values <0.01 .

Results and discussion

Synteny with divergence of the ancestrally tetrasomic chromosome

The *L. passim* nuclear genome was assembled to 31 chromosome-level scaffolds, along with the mitochondrial (kinetoplast) maxi- and minicircles typical of trypanosomatids, as described previously [40]. The genome assembly displayed a high degree of synteny with related species in the Leishmaniinae subfamily (Fig. 1), supporting the phylogenetic placement of *L. passim* within the *Crithidia/Leptomonas/Lotmaria* clade (Fig. 2A). Synteny was highest with the close relative *Leptomonas pyrrhocoris*, for which the *L. passim* genome was 97% covered by 49 synteny blocks, including 62 sets of at least 20 collinear genes (Figs. 1 and 2B). However, synteny was also strong with *Crithidia bombi* (87% across 71 blocks) and *C. fasciculata* (92% across 57 blocks), and slightly lower but with fewer blocks for the chromosome-level assemblies of *Leishmania major* (86%, 41 blocks) and *L. braziliensis* (86%, 48 blocks) (Fig. 2B). In four cases (Chromosomes 1, 2, 4, and 13), *L. passim* chromosomes displayed synteny to multiple full chromosomes of *L. major* (Fig. 1A). For example, *L. passim* Chromosome 1 incorporated synteny blocks spanning *L. major* Chromosomes 36 and 3, similar to the fusion of *L. major* Chromosomes 36 and 20 in *L. mexicana* [16].

Compared to the rest of the genome, the paralogous *L. passim* Chromosomes 5 and 6 showed substantially lower levels of synteny with homologous chromosomes in most of the trypanosomatid relatives we examined, with the exception of *L. pyrrhocoris* (Fig. 2B). Only around half the combined length of these chromosomes showed synteny with *C. bombi* and *C. fasciculata*, and no synteny was detected with *C. expoeiki* or either species of *Leishmania*, contrasting with the $>80\%$ synteny of the genome overall (of which Chromosomes 5 and 6 constitute $\sim 12\%$) for each of these comparator species (Fig. 2B). However, each of Chromosomes 5 and 6 was syntenic with the tetrasomic scaffolds 12, 29, and 32 of the *L. pyrrhocoris* assembly (Fig. 3) [9, 14], suggesting that these three scaffolds (ordered 32, 12, 29) comprise a single *L. pyrrhocoris* chromosome. These *L. pyrrhocoris* scaffolds were in turn syntenic with the ancestrally tetrasomic Chromosome 31 of *L. major*, strongly suggesting that these *L. passim* chromosomes share ancestry with the *Leishmania* Chromosome 31. This synteny, however, was not inferred when comparing *L. passim* with *L. major* directly (Fig. 1A). These findings suggest that Chromosomes 5 and 6 have substantially diverged from their ancestral state in *L. passim* and compared to the equivalent genomic regions of other monoxenous species.

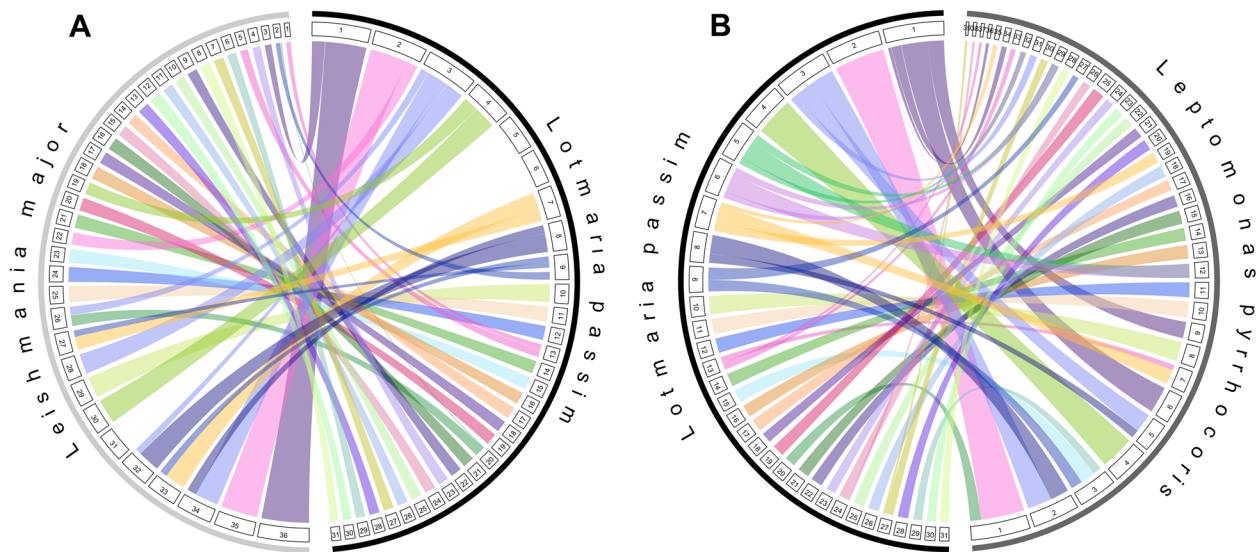


Fig. 1 Synteny of *L. passim* chromosomes with trypanosomatid relatives *Leishmania major* and *Leptomonas pyrrocoris*. Labeled bands depict scaffolds of *L. passim* and (A) *L. major* or (B) *L. pyrrocoris*. Arcs connect syntenic regions. **A** The paralogous *L. passim* Chromosomes 5 and 6 and the tetrasomic *L. major* Chromosome 31 constitute the only chromosome-scale synteny gaps. **B** All major genomic regions are shared between *L. passim* and *L. pyrrocoris*, with the region syntenic with paralogous Chromosomes 5 and 6 split across 3 *L. pyrrocoris* scaffolds

The combination of a high quality assembly and high nucleotide divergence between the two haplotype pairs (88.8% sequence identity) allowed us to distinguish two separate disomic chromosomes in *L. passim* (Fig. 4). Our finding was supported by both median sequencing depth—which was within 20% of the overall chromosome-level median for each of the two scaffolds—and variant calling, which was consistent with a 50:50 distribution of variants at polymorphic loci [40]. Although this contrasts with the tetrasomy concluded for the other species [14, 16], divergence of these chromosomes is consistent with the high nucleotide diversity of the tetrasomic genome scaffolds within populations of *Crithidia*, *Leptomonas*, and *Leishmania* relatives [14]. Additional haplotype-level sequences of this tetrasomic region in other trypanosomatids are needed to determine whether its assortment into two chromosome pairs is unique to *L. passim*. However, analysis of the *L. major* Friedlin genome found evidence for only 239 polymorphic sites on Chromosome 31 [66], suggesting >99.98% identity across haplotypes over this ~1.5 Mb region.

The validity of the assembled chromosomes was supported by the presence of telomeric repeats at their termini. Most of the telomeres contained the canonical tandem arrays of “CCCTAA” commonly used to distinguish trypanosomatid telomeres [46] (Supplementary Fig. 1). However, other tandem repeats that did not include this motif were found as well, accounting for 45% of the total number (Supplementary Table S3). For example, Chromosome 7 began with over only 2 repeats

of “CCCTAA” (Supplementary Fig. 1) followed by 3 Kb of repeating “ACACCCGT”.

Although slight variations in CCCTAA-containing repeats have been found across species and among chromosomes of *Leishmania* species [67, 68], our assembly suggests a diversity of repeat structures that has not been resolved by the use of CCCTAA-based probes [46], and will require additional chromosome-level assemblies to better understand.

Structural patterns of gene arrangement and transcription *Genes are arranged in polycistronic clusters*

Our annotation of genes indicated the presence of long clusters of genes on the same strand that in some cases spanned entire chromosomes (Fig. 4, inner ring shading), consistent with the previously described organization of trypanosomatid genes in long polycistronic units [11]. Adjacent gene pairs had a 97% chance of being located on the same strand; genes within 1 Kb of one another had a >90% chance of sharing a strand, and those within 10Kb a >80% chance (Supplementary Fig. 2). The strand-wise arrangement of genes was conspicuously strong on the paralogous Chromosomes 5 and 6 (Fig. 4, inner ring shading), where essentially all (96%) of the annotated regions (including intact protein-coding genes as well as non-coding RNAs and predicted pseudogenes) were on the negative strand, consistent with annotations of the tetrasomic chromosomes with shared ancestry in other trypanosomatids [14, 16].

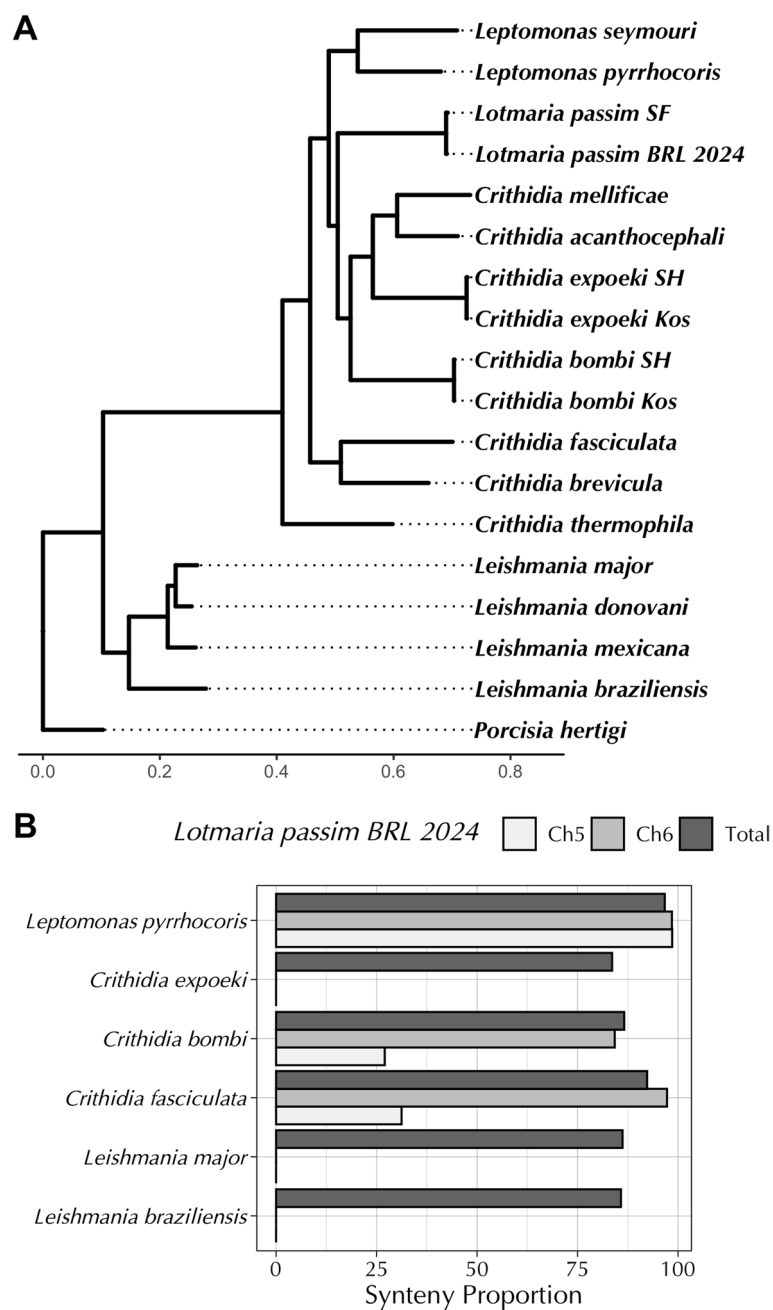


Fig. 2 Phylogenetic placement and proportions of the *L. passim* genome and paralogous Chromosomes 5 and 6 syntenic with genomes of selected relatives. **A** Phylogenomic tree depicting relationships between *L. passim*, *L. pyrrhocoris*, and other trypanosomatids in the Leishmaniinae subfamily. The tree was inferred from protein sequences of 2530 single-copy orthologs. Our assembly is labeled 'BRL (2024)'. The position of the 'SF' strain used for the first draft genome assembly is also shown [41]. For *C. bombi* and *C. expoeki*, nodes are shown for both the original published annotations by Schmid-Hempel and colleagues ('SH' [89]) and independent annotations of the same assemblies created in a review of trypanosomatid phylogenomics by Kostygov and colleagues ('Kos' [8]). All nodes had support values of 1 based on 1,000 resamples. **B** Synteny of paralogous chromosomes and genome overall. Y-axis shows species analyzed; X-axis represents proportion of the *L. passim* genomic of chromosome region covered by synteny blocks. Shading of bars corresponds to the region quantified. "Total" indicates the entire nuclear genome

Ancestrally tetrasomic, paralogous Chromosomes 5 and 6 are enriched in genes upregulated during infection

We used publicly available *L. passim* transcriptomic data [32] in combination with our new chromosome-level genome assembly and its annotation to re-examine

structural and functional patterns of gene expression during infection of honey bees. Roughly two-thirds of annotated genes were differentially expressed in bees vs. log-phase cell cultures at each of the four time points post-inoculation (Fig. 5A). The paralogous Chromosomes

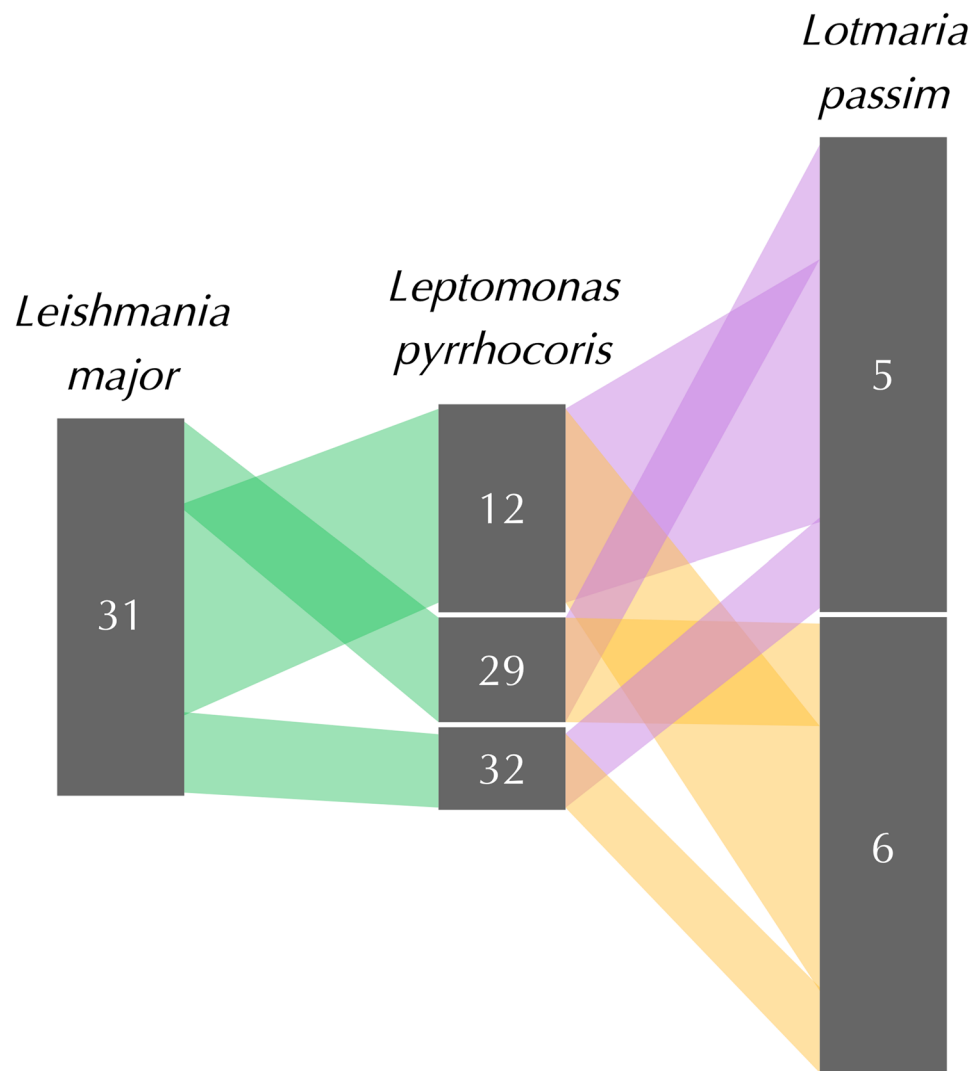


Fig. 3 Three-way synteny between *L. passim* Chromosomes 5 and 6; *L. pyrrocoris* tetrasomic scaffolds 12, 29, and 32; and *L. major* tetrasomic Chromosome 31. The *L. passim* Chromosomes 5 and 6 are syntenic with the same *L. pyrrocoris* scaffolds that showed synteny with *L. major* Chromosome 31. This strongly suggests shared ancestry between the corresponding chromosomes of *L. passim* and *L. major*

5 and 6 were conspicuously enriched in genes upregulated in the bee gut environment, with higher probabilities of genes being upregulated and lower probabilities of genes being downregulated (Fig. 5A). Across all time points, genes on these two chromosomes were more than twice as likely to be upregulated by greater than two-fold relative to genes on other chromosomes (Fig. 5B). However, genes on the duplicated chromosomes were expressed at more than two-fold lower levels (i.e., fewer read counts) than those on other chromosomes (Supplementary Fig. 3; Supplementary Fig. 4), thereby compensating for the doubling of copy numbers resulting from chromosome duplication. A similar pattern, in which genes on the ancestrally tetrasomic chromosome are collectively expressed at similar levels to those on disomic chromosomes despite having twice as many copies, was

likewise reported for the *Leishmania* Chromosome 31 [14]. The paralogous chromosomes were also enriched in predicted pseudogenes, with a > 50% increase in proportions of this feature type relative to the rest of the genome (Supplementary Fig. 5).

Transcript abundances show traces of polycistronic transcription

We quantified spatial patterns of gene expression within chromosomes and their relationship to strand-wise gene arrays, which are canonically transcribed as a single polycistronic unit [11]. We found weak (correlation coefficient 19–24%) but highly significant positive correlations between expression levels of adjacent genes in each sample group (e.g., $\chi^2_1 = 671$, $P < 0.001$ for cell cultures with low-count genes excluded). However,

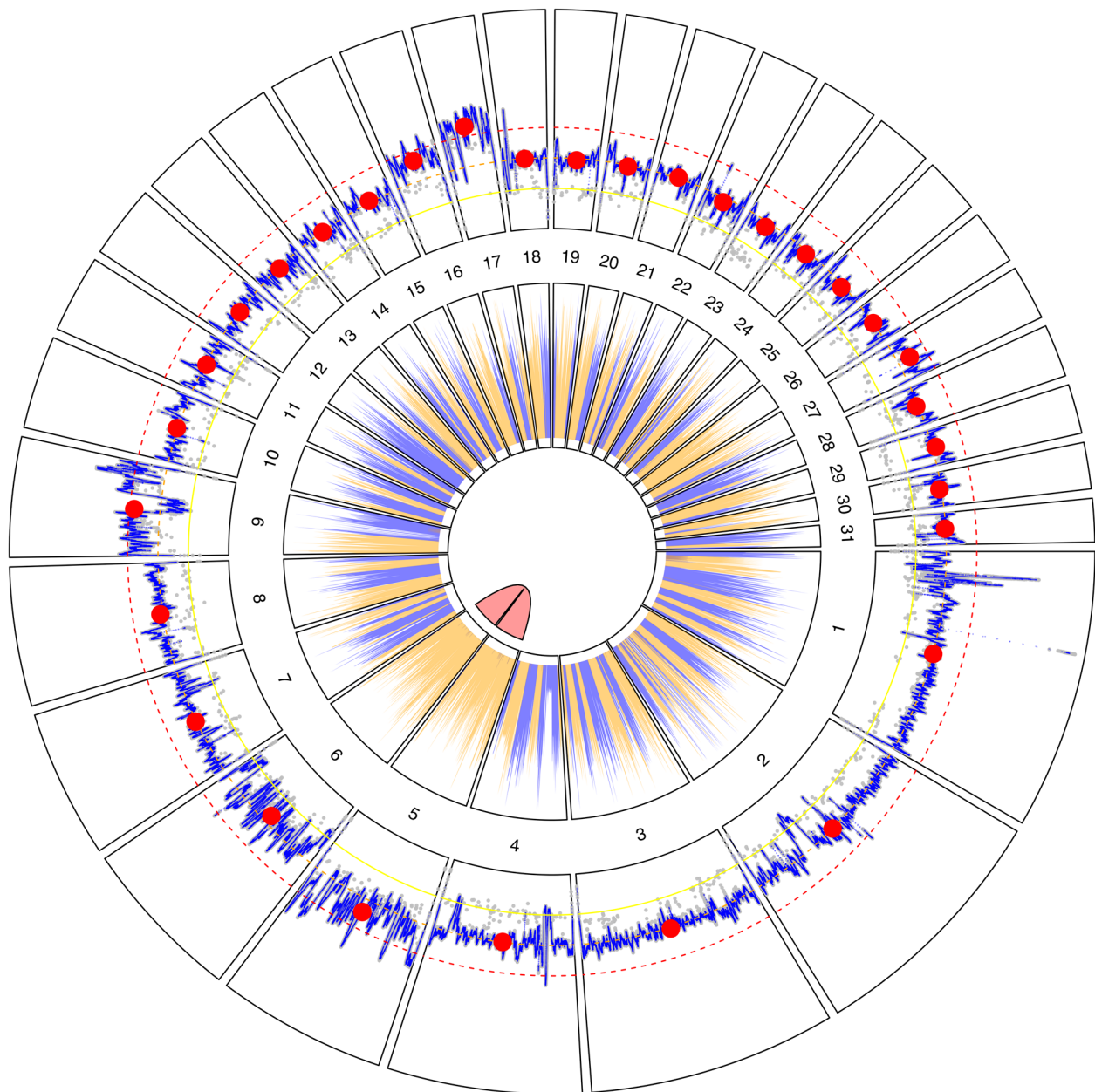


Fig. 4 Circular representation of *Lotmaria passim* BRL (2024) genome assembly. Radial segments correspond to the 31 nuclear chromosomes, numbered by size from largest to smallest. Outer ring: sequencing depth. Gray points show base-level read depth, subsampled at 100 bp intervals. Blue line trace represents 1 Kb moving average. Large red circle represents the chromosome-level median. Faint scatter of gray points with roughly half the read depth of the chromosome overall suggests heterozygous sites. Concentric yellow, orange, and red lines represent 50%, 100%, and 150% of the median chromosome-level depth; this corresponds to expected depths for monosomic, disomic, and trisomic chromosomes, respectively. The paralogous Chromosomes 5 and 6 appear as disomic chromosomes with distinct coverage traces. Inner ring: relative density of coding regions (percent of each 10 Kb interval covered by predicted exons), colored by strand (blue: positive, orange: negative), suggesting long stretches of polycistronic genes on the same strand and strand-biased gene arrangement on Chromosomes 5 and 6. Opaque red link between Chromosomes 5 and 6 indicates the similarity between these two regions

the correlation between neighbors was stronger for neighboring genes on the same versus opposite strands of the chromosome, as evidenced by the significance of the neighbor expression $\times$ shared strand interaction term (e.g., $\chi^2_1 = 23.6$, $P < 0.001$ for cell cultures, Fig. 6). The fixed-effects model explained 5–10% of the variance

in each sample group, suggesting a detectable signal of polycistronic transcription despite considerable post-transcriptional regulation of mRNA levels; this observation might be at least partly attributed to the origin of a portion of sequencing reads from precursor mRNAs. Same-stranded arrays of >10 consecutive genes

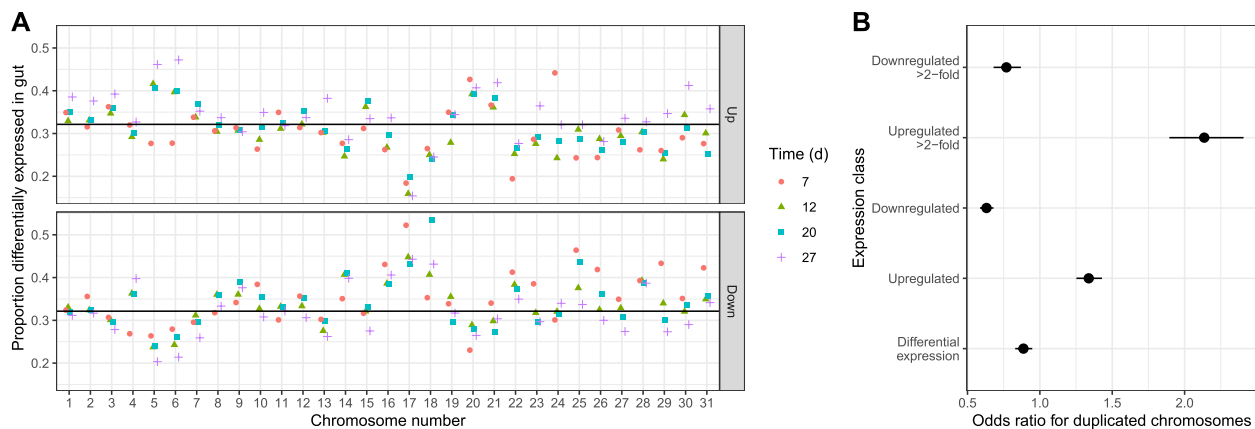


Fig. 5 Paralogous Chromosomes 5 and 6 are enriched in genes upregulated during infection of honey bees. **A** Proportions of genes significantly ($P < 0.05$) up- and down-regulated for each chromosome and time point post-inoculation. Horizontal lines indicate median proportions. Shapes and colors represent the four times of sampling, in days post-inoculation. **B** Exponentiated model coefficients for proportions of genes significantly ($P < 0.05$) and strongly ($P < 0.05$ and > 2 -fold change) up- and down-regulated in bee gut vs. culture environments. X-axis shows exponentiated coefficients and 95% confidence intervals for genes on Chromosomes 5 and 6 versus the rest of the genome

explained 8–16% of variation in expression levels when modeled as a random effect, depending on whether genes with low read counts were excluded (Supplementary Fig. 6).

We also found positive correlations between differential expression of neighboring genes in bee guts vs. cell culture, although these were less pronounced than for absolute expression, and no significant difference was evident between gene pairs on the same vs. opposing strands (i.e., $P > 0.05$ for neighbor $\log_2$ -fold change by shared strand interaction term, Supplementary Fig. 7, Supplementary Table S4). The proportion of variance explained by the fixed effects model (comprising $\log_2$ -fold change of the neighboring gene, whether neighbors were on the same strand, and their interaction) was $< 1\%$ for every time point, indicating that differential expression of neighboring genes is only weakly interdependent.

Several factors could contribute to the relatively low correlations in transcription between genes within polycistronic gene clusters in our analysis. Technically, it is difficult to define true transcription start sites. In trypanosomatids, such sites lack promoter motifs and can include multiple potential start positions that span a 50–100 bp region epigenetically marked by histone acetylation [69]. In the absence of epigenetic evidence, our gene clusters were delineated based on the strand switches of genes in our annotation and may not correspond to the actual transcriptional units. Trypanosomatid mRNA abundance is subject to extensive post-transcriptional regulation, most notably by RNA-binding proteins that mainly bind to 3' untranslated regions and modulate mRNA stability [11]; this compensates for the lack of gene-level transcriptional control. Indeed, the fact that some of our sequencing reads likely

originate from precursor (rather than exclusively mature) mRNAs not yet subjected to these regulatory mechanisms could have resulted in generous estimates of inter-genic correlation in our analyses.

Functional analysis of transcription

We used the annotation of our new genome assembly [40], which identified roughly 2,000 more protein-coding genes than the annotation of the draft genome assembly [41] used previously for RNAseq analysis of *L. passim* [32], to analyze functional patterns of differential gene expression in the gut vs. cell culture environment using Gene Set Enrichment Analyses. This method, which characterizes the rank-based distribution of changes in expression level for the set of genes associated with each function-related term [58], offers a more quantitatively nuanced alternative to the binomial enrichment tests for overrepresentation used by the dataset's original authors [32], incorporating information about the magnitude of differential expression rather than relying on arbitrary distinctions between genes that are differentially expressed or not. We also conducted a pathway analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [61], which contains detailed information on metabolic and other pathways, including those specifically related to trypanosomatid infection.

Upregulation of peptidase activity and cell adhesion; downregulation of carbohydrate metabolism, protein synthesis, and cell motility during infection

Gene ontology (GO) term enrichment analyses were generally consistent across different time points post-inoculation, with many of the same terms occurring repeatedly among the most highly enriched gene sets (Fig. 7). Most of the strong enrichment scores were

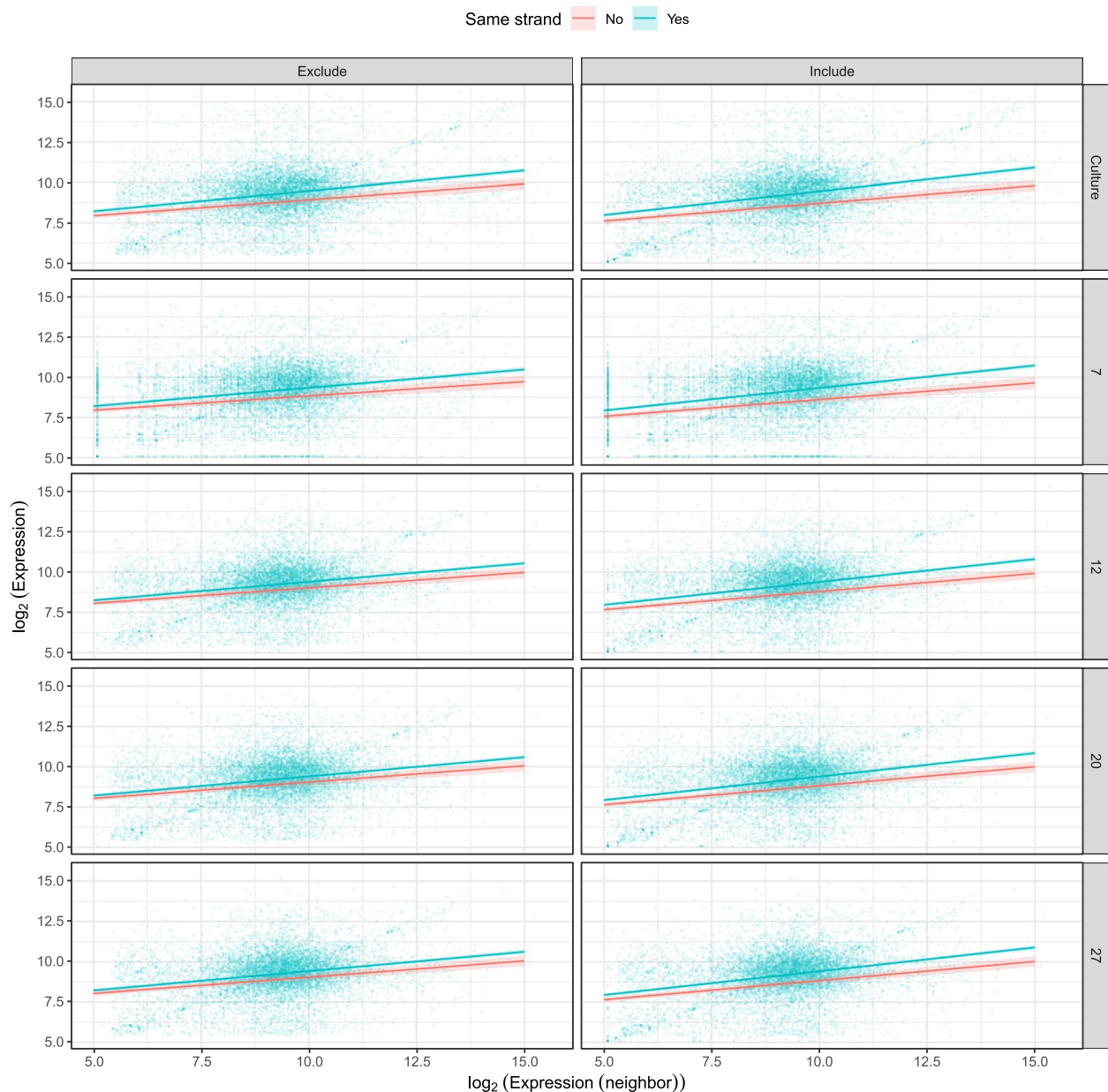


Fig. 6 Expression levels of adjacent genes are correlated, with stronger correlations between neighboring genes located on the same strand of the chromosome. Y- and X-axes show log-transformed expression levels (transcripts per million) of each pair of neighboring genes. Trendlines show correlations with 95% CI shaded bands, in red for neighbors on opposite strands and blue for neighbors on the same strand. Panels in rows indicate sample groups (cultured cells and cells from bee guts at different time points post-inoculation). Panels in columns show data with and without low-count (i.e., low expression level) genes excluded

associated with down-regulated gene sets. There was a general decrease in expression of gene sets related to carbohydrate metabolism, ribosomal protein synthesis, and DNA replication, consistent with adaptations to a low-nutrient environment in the bee hindgut. There was also a decrease in expression of genes involved in responses to oxidative stress and detoxification (see Supplementary Data S4). These gene sets were largely overlapping, with leading-edge genes including catalases, glutathione and cytochrome peroxidases, and peroxiredoxins. Their

upregulation in the culture environment could reflect responses to either oxidative or chemical stress. This could result from culturing under ambient oxygen levels as opposed to the relatively low oxygen levels in the gut [37]. Alternatively, it could reflect the fact that cultures were grown in antibiotic-enriched media, which was anecdotally reported to ‘massively impede’ growth of the related *C. bombi* and prevent establishment of new cell lines [54]. Upregulated gene sets were dominated by terms related to peptidase activity and amino acid

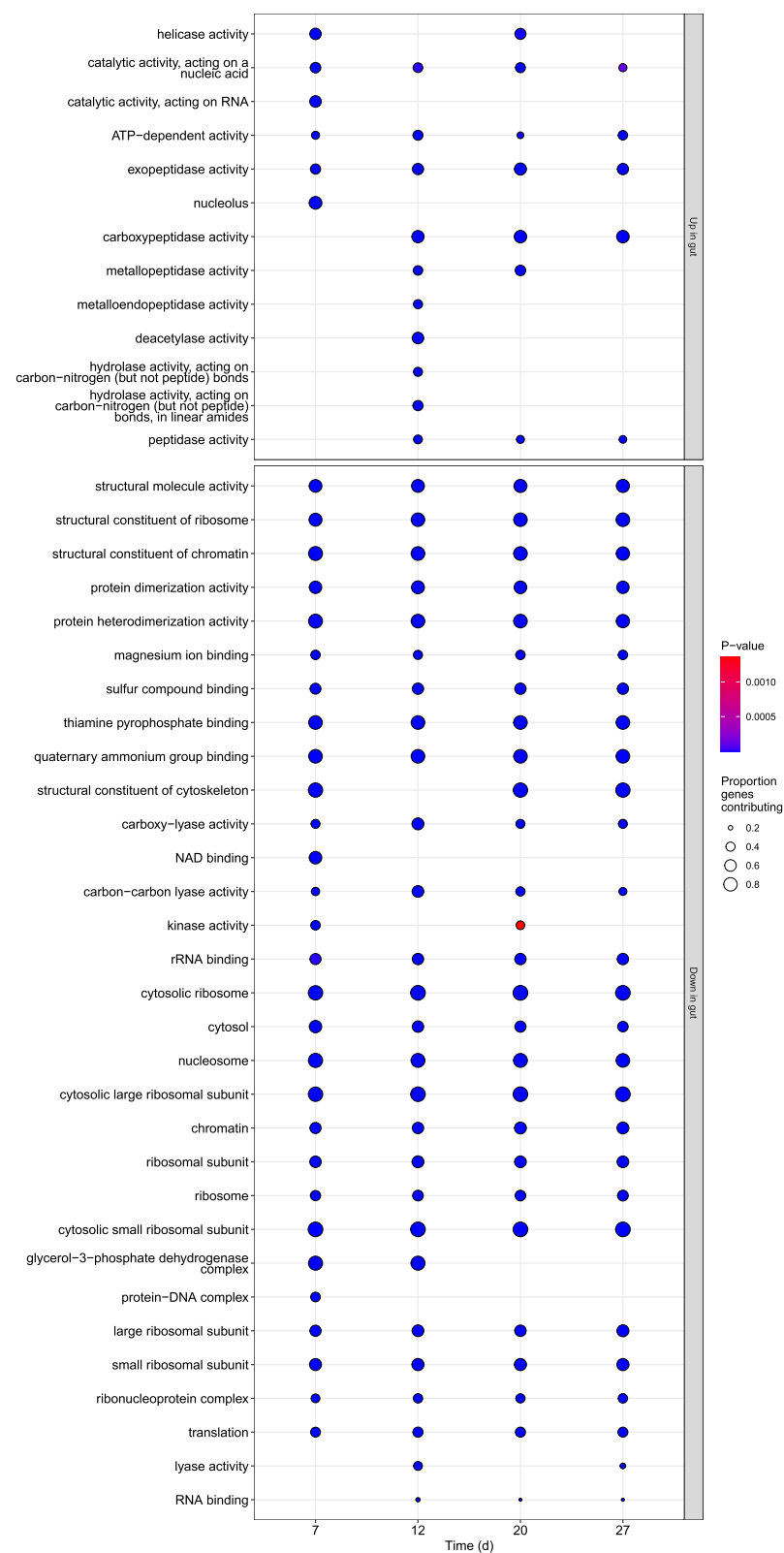


Fig. 7 Gene Ontology gene set enrichment for *L. passim* expression levels in honey bee gut samples at 4 time points post-inoculation relative to cultured cells. Y-axis shows the term description. Panels show up- and down-regulated gene sets. Shading is proportional to the p-value for the test and size to the proportion of annotated genes that contribute to the up- or down-regulation (i.e., 'leading edge' genes). Only terms with the top 30 enrichment scores for each time point are shown

metabolism has suggested a shift from carbohydrate- to amino acid-based substrates, similar to changes reported in other trypanosomatids in the insect gut vs. rich culture media [70] and conclusions from the initial publication describing this dataset [32].

We also found downregulation of gene sets related to microtubules, the cytoskeleton, and cell projections, with several flagellar proteins among the genes contributing to the downregulation (see Supplementary Data S4). This pattern was consistent with observations of reduced flagellum length and cell motility in attached haptomonad morphotypes carpeting the bee gut epithelium, as opposed to the freely swimming promastigotes found in log-phase cultures [22, 23]. This was countered by upregulation of gene sets related to protein and vesicle-mediated transport, macromolecule localization, and the endomembrane system (Supplementary Data S4), suggesting reorganization of cellular structure and intracellular machinery as cells proliferated along the gut wall (Fig. 7).

Our GO reanalysis differs from that of the original authors [32] in data inputs and methodological approach, but yields similar conclusions. Both analyses indicate increased expression of peptidases; but downregulation of genes involved in energy metabolism, protein synthesis, oxidative stress, and flagellar function. One gene set not noted by the original authors was the ABC transporters, which showed upregulation at the final time point (Supplementary Data S4). Leading edge genes included multiple paralogs of pentamidine resistance protein 1, ABC subfamily B and C proteins, and mitochondrial multidrug resistance proteins. This induction of expression could result from increased exposure to host-produced antimicrobial peptides and chemicals from plants and the environment as bees become foragers later in life [71]. Our analysis additionally identified upregulation of the cell adhesion gene set, with leading edge genes comprising five GP63/leishmanolysin paralogs, which we discuss further in the context of the KEGG analysis below.

KEGG enrichment suggests upregulation of glycan and amino acid synthesis, downregulation of flagellar genes in gut environment

Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was concordant with the GO analysis, indicating downregulation of gene sets related to protein synthesis, energy metabolism and glutathione-based antioxidant metabolism, and flagellar and motor proteins. This was countered by upregulation of gene sets related to amino acid metabolism (arginine biosynthesis; 2-oxocarboxylic acid metabolism; and alanine, aspartate, and glutamate metabolism; Fig. 8). Upregulation of the autophagy-associated gene set at the first time point (7 d post-inoculation) was further

suggestive of a nutrient-limited environment in the bee gut; this term was not highlighted by the original study's GO term analysis [32]. We also found upregulation of the gene set encoding proteins for variants of GP63/leishmanolysin—a glycoprotein and peptidase involved in cell adhesion and cleavage of host effector proteins. The N-glycan biosynthesis-related gene set was likewise upregulated, which was not reported previously [32]. Upregulated genes in this set included several glycosyltransferases, which catalyze the attachment and removal of sugar and acetylglucosamine groups from proteins (Supplementary Fig. 8). Both GP63 and N-glycan biosynthesis gene sets are likely important for adherence to and interaction with host tissues or gut biofilms [72, 73].

Attachment to the insect gut epithelium is a process common to most of the trypanosomatids, with cell surface proteins playing a key role in the process [72]. In *C. fasciculata*, which like *L. passim* attaches to the hindgut wall and assumes a truncate morphology with a reduced flagellum, similar upregulation of GP63 proteins and genes involved in cell adhesion was observed during infection of mosquitoes as well as in adherent cultured cells [73]. Given that the attachment process generally occurs on the time scale of hours rather than days [72], closer attention in the period immediately post-inoculation could provide greater insight into how *L. passim* gene expression changes during initiation of infection.

Overall, the functional patterns of differential expression bear strong similarity to those reported in other trypanosomatids during both insect and mammal infection. Transcriptomes of *Herpetomonas muscarum* in *Drosophila* relative to cell cultures likewise indicated a pattern of increased amino acid utilization and autophagy [74]. Like the changes in *C. fasciculata* [73], they also showed reduced expression of flagellum-related genes and increased expression of multiple GP63 peptidases as well as other proteins associated with the cell surface [74]. In addition, post-inoculation upregulation of genes involved in DNA replication and repair in *H. muscarum* [74] was concordant with the upregulation of gene sets annotated for nucleic acid catalytic and helicase activities in our GO term enrichment analysis.

Beyond the insect gut, several aspects of *L. passim* gene expression in bees resembled those found in *Leishmania* species in the transition from the insect-associated promastigote form (in cell culture) to the amastigote form associated with vertebrate blood cells. These include down-regulation of gene sets associated with growth, carbon metabolism and flagellar motility; and upregulation of gene sets encoding peptidases and cell surface proteins [16, 75]. In light of these similarities, it is understandable how the same genes and related regions of the genome, such as the glycoprotein anchor and other

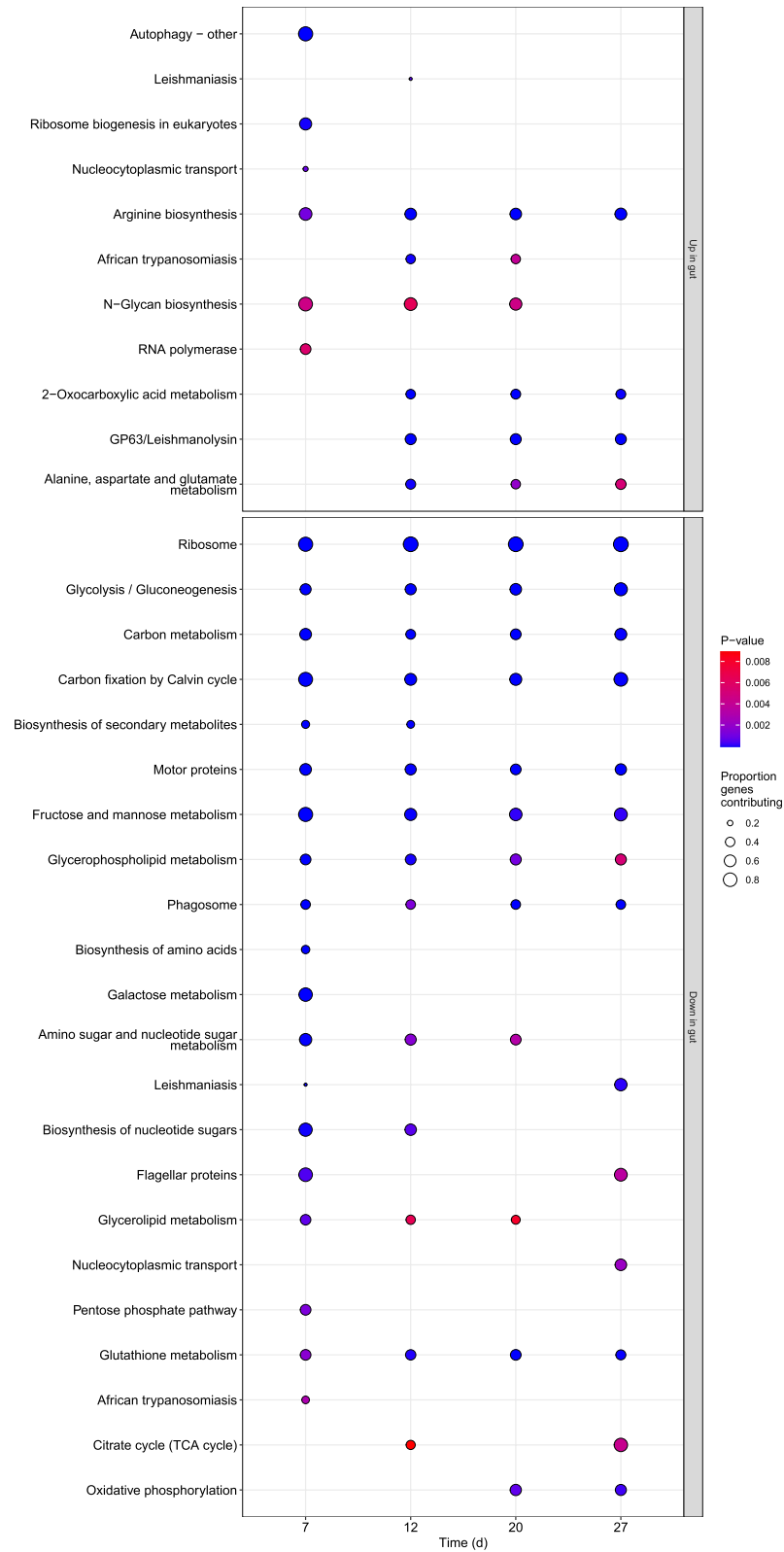


Fig. 8 KEGG pathway gene set enrichment for *L. passim* expression levels in honey bee gut samples at 4 time points post-inoculation relative to cultured cells. Y-axis shows the term description. Panels show up- and down-regulated gene sets. Shading is proportional to the p-value for the test and size to the proportion of annotated genes that contribute to the up- or down-regulation (i.e., ‘leading edge’ genes)

surface protein-associated genes that are enriched on the ancestrally tetrasomic chromosome [14], could be important for both insect and vertebrate infection, and why this supernumerary region of the genome was enriched in genes upregulated during both *Leishmania* infection of blood cells [16] and *L. passim* infection of the bee gut described here. Maintaining and diversifying additional copies of such niche-specific, surface protein-encoding genes could enable parasites to continually evade recognition and subvert attack by current and novel hosts across evolutionary time.

Paralogous chromosomes contribute to upregulation of peptidase activity, amino acid synthesis, and ABC transporters

To explore the specific contributions of the paralogous Chromosomes 5 and 6 to differential gene expression in bees, we looked for gene sets upregulated during infection with at least 20% of their leading edge genes on Chromosomes 5 and 6 (i.e., more than expected given their combined ~12% of the genome size). There were 40 GO term-time point combinations that met these criteria. All but three of the terms involved some aspect of amino acid metabolism, including arginine and glutamine metabolic or biosynthetic processes and peptidase or carboxypeptidase activity at every time point; a consistent set of nine glutamyl carboxypeptidases on the paralogous chromosomes were among the leading edge genes. Eight KEGG pathway-timepoint combinations also met this criteria, of which seven included at least three leading-edge genes on the paralogous chromosomes. However, these comprised only two terms related to amino acid metabolism (2-oxocarboxylic acid metabolism and arginine biosynthesis); for each, the leading-edge genes matched the glutamyl carboxypeptidases highlighted by the GO analysis. These sets highlight the contributions of the paralogous chromosomes to the upregulation of amino acid metabolism, and specifically peptidase activity, that emerged from both GO and KEGG enrichment analyses.

The highest proportion of leading-edge genes on Chromosomes 5 and 6 (18/34, 53%) was found for the ABC-type (ATP-binding cassette) transporters upregulated at 27 d. These genes included four paralogs annotated as pentamidine resistance protein 1, four as putative mitochondrial p-glycoprotein or multidrug resistance protein E, six as ABC protein subfamily C, and two as subfamily D. These proteins have been implicated in drug resistance in *Leishmania* [76]. Their upregulation in the bee gut could enable parasite tolerance of the diverse phytochemicals found in nectar and pollen [77], which occur at concentrations sufficient to inhibit growth of *Leishmania* [78] and are proposed to have shaped the evolution of detoxification capacities in bees [79].

We also inspected the annotations of the remaining genes upregulated during infection (adjusted $P < 0.05$). Of 1,631 gene-timepoint combinations, only 23% were defined by the GO annotation and 12% by the KEGG; 74% had none. Peptidases were again prominent, including 34 cases of upregulation of glutamyl carboxypeptidases, eight of zinc metallopeptidases, and four of calpain family cysteine proteases. There were also 44 instances of upregulation of ABC-type transporters, 10 of zinc transporters, and 8 of carnitine transporters.

There were additional upregulated genes that could be important for attachment to the bee gut, remodeling of the cell surface, and interaction with host immune pathways. Putative modulators of cell adhesion included the calpains noted above – which play a role in *Trypanosoma cruzi* attachment to the tsetse fly midgut [80]—eight instances of upregulation of glycoprotease 2 genes, and three of GP63 genes. There were also five instances of upregulated phosphoglycan galactosyltransferases likely involved in lipophosphoglycan (LPG) synthesis and eight of integrin alpha chain protein. Upregulation of genes involved in vesicle-mediated transport—including endoplasmic reticulum vesicle transporters, BAR- and GOLD-domaining proteins, and Ras-like small GTP-ases – suggested an additional role of the paralogous chromosomes in transport of these mediators of attachment to the cell surface, while upregulation of genes involved in sterol synthesis (mevalonate kinase, sterol desaturase, and squalene synthase) could support remodeling of the cell membrane in the attached state. Also intriguing were six cases of upregulation of prostaglandin synthases, which could modulate host immune pathways. Collectively, these findings suggest that genes on the paralogous chromosomes modulate primary metabolism and nutrient acquisition, interactions with hosts, and tolerance of chemicals in the gut milieu.

Expression of leishmaniasis and trypanosomiasis infection-related pathway gene sets varies across time points

The most apparent fluctuations in differential expression across time points were found for gene sets annotated as components of the leishmaniasis and trypanosomiasis pathways, each of which appeared among the top up- or downregulated sets at different time points (Fig. 8). Both terms describe genes involved in trypanosomatid infection of mammals, consisting mostly of host immune pathways along with a few parasite-derived virulence factors and immunomodulators. The leishmaniasis pathway gene set was down-regulated at 7 d, upregulated at 12 d, then downregulated again at 20 d post-inoculation. This reflected strong downregulation of cysteine peptidase B (CPB) at 7 d and 20 d and downregulation of elongation factor 1-alpha (EEF1a) at 20 d, which was countered by strong upregulation of GP63-family genes from 12 d

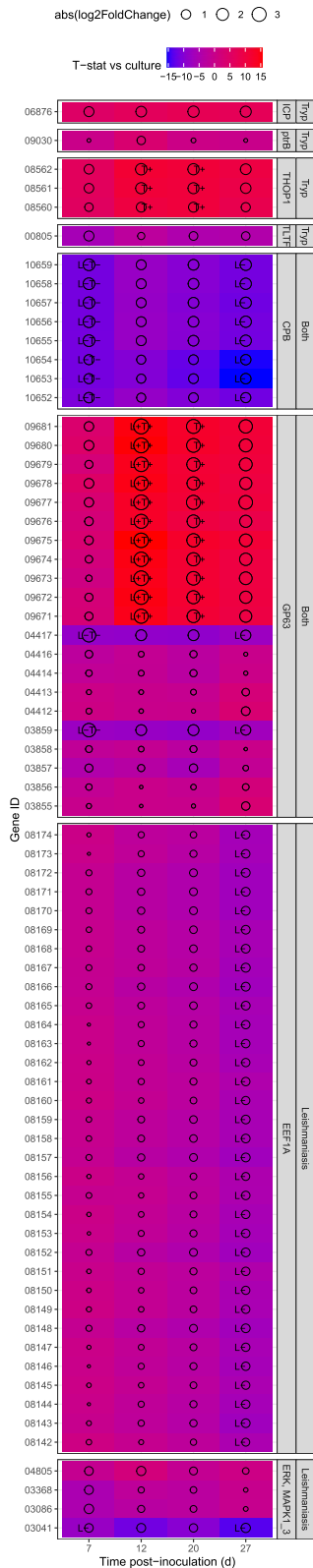


Fig. 9 Differential expression of genes annotated as constituents of the 'African trypanosomiasis' (Tryp) and 'Leishmaniasis' KEGG pathways. Shading corresponds to T-statistic for differential expression for bee gut vs. cultured cells at each time point post-inoculation. Open circles are sized in proportion to the log₂-fold change. Y-axis numbers correspond to paralogs within each gene type, numbered consecutively according to their position in the genome. Panels labeled 'both' correspond to genes included in both the trypanosomiasis and leishmaniasis pathways. Letters and signs within the plot indicate genes most strongly contributing to up- ("+") or downregulation ("-") of the leishmaniasis ("L") and trypanosomiasis ("T") pathways at each time point. Abbreviations: ICP: inhibitor of cysteine peptidase; ptrB: protease (Oligopeptidase) B; THOP: thimet oligopeptidase ("Tropolysin"); TLTF: T lymphocyte-triggering factor; CPB: cysteine peptidase B; GP63: major surface glycoprotein 63 ("Leishmanolysin"); EEF1A: elongation factor 1-alpha; ERK: extracellular signal-regulated kinase, MAPK: mitogen-activated protein kinase

onwards (Fig. 9). The trypanosomiasis pathway gene set was downregulated at 7 d—again reflecting downregulation of cysteine peptidase B (CPB)—but upregulated at 12 and 20 d, reflecting increased expression of multiple paralogs of GP63 and thimet oligopeptidase (THOP1/tropolysin) (Fig. 9).

GP63 is noted for providing resistance to complement-mediated lysis in the bloodstream as well as to hydrolytic enzymes in the sand fly gut [81], whereas cysteine peptidase inhibits inflammatory signaling pathways [82]—including the JAK-STAT pathway important for clearing infection with the trypanosomatid *H. muscarum* in *Drosophila* [83]. Although both represent large protein families, this suggests that these genes could likewise protect parasites from host (bee) defenses or suppress their expression in response to infection. Indeed, very few changes in host gene expression occurred in *L. passim*-inoculated honey bees [32]; only modest and transient upregulation of immune genes occurred in honey bees inoculated with *C. mellificae* [84]; and most bumble bee immune genes showed little induction following inoculation with *C. bombi* [85], with the most infectious strains eliciting the weakest expression of antimicrobial peptides [86].

Uncertainty about expression of individual paralogs and abundance of protein products

Scrutiny of genes involved in the leishmaniasis and trypanosomiasis pathways reflected both the structure of the genome and the limitations of the RNAseq analysis. Strings of paralogs generally appeared as tandemly arrayed genes along the chromosome, including 2 sets of 5 paralogs of GP63 on the Chromosomes 5 and 6 and over 30 consecutive paralogs of EEF1A on Chromosome 18 (Fig. 9). Due to the sequence similarity among paralogs, sequencing reads could not be unambiguously assigned. These reads were distributed fractionally across each gene

to which they aligned, resulting in read counts that were identical (or nearly so) among many spatially clustered paralogs, and non-independence of differential expression estimates for each one. This limitation in precision of mapping also likely inflated our estimates of correlations between transcript levels of neighboring genes (Fig. 6).

An additional limitation is our reliance on transcriptional data as the sole metric of gene expression. Abundance of mRNA correlates imperfectly with concentration of the corresponding proteins, and in *Leishmania* there is a dynamic relationship between the two [87]. Characterization of the proteome—recently described for *L. passim* in cell culture [88]—and its changes during different phases of infection would provide further insights into regulation of these ultimate gene products and how it compares with patterns of transcription.

Conclusions

Our analysis leverages one of the first chromosome-level assemblies of an insect-specific trypanosomatid parasite of bees to illuminate the structural singularities of the genome and how they relate to gene expression, host infection, and parasite evolution. By linking the differentiation of an ancestrally tetrasomic region (i.e., Chromosome 31 of *L. major*) into two distinct, disomic chromosomes (i.e., Chromosomes 5 and 6 of *L. passim*) with the presence of genes upregulated in the honey bee host, our findings suggest a potential advantage of this salient genomic feature. Our transcriptomic analysis highlights shared patterns of gene expression across diverse trypanosomatids during host infection, as well as similarities between the functional changes that occur during infection of insects and vertebrates. The paralogous chromosomes, including genes for surface proteins that directly interface with hosts, appear to have disproportionate importance for establishment in both types of host. Hence, both elevated copy number and divergence of these chromosomes and their genes was likely advantageous in colonizing the diverse ecological niches occupied by trypanosomatids, including the hindgut of honey bees.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-12082-y>.

Supplementary Data S1. RNAseq read counts used for transcriptome analysis.

Supplementary Data S2. GO assignments made by Pannzer2.

Supplementary Data S3. KEGG assignments made by BlastKoala.

Supplementary Data S4. GSEA results from the GO analysis.

Supplementary Data S5. GSEA results from the KEGG analysis.

Supplementary Information: Supplementary Tables, Figures, Methods, and description of data files.

Acknowledgements

The authors thank Amanda Albanaz for proteome annotations of *Crithidia mellificae*, *Crithidia bombi*, and *Crithidia expoeki*; and two anonymous reviewers for their constructive feedback on the original manuscript.

Authors' contributions

AN, JDE, and ECPY conceived the study. AN and ECPY analyzed the data with guidance from AB. ECPY wrote the first draft of the manuscript. All authors revised the manuscript.

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 the Czech Science Foundation grant 23-07695S to AB.

Data availability

The *L. passim* BRL (2024) Hi-C and HiFi raw reads are available on NCBI GenBank (accession ID: SRX22798691 and SRX22798690); the assembled chromosomes (accession ID: GCA_037349495.1), maxicircle, and minicircles are listed under BioProject PRJNA1049372. The annotated genome is available on FigShare as a supplement to our previous article documenting the assembly [40] at <https://doi.org/10.25387/g3.27121227>. RNA sequence data were used from NCBI BioProject PRJNA587465. The gene counts, GO terms assigned by Pannzer2, and KEGG terms assigned by BlastKOALA are provided as supplementary materials.

Declarations

Ethics approval and consent to participate

N/A

Consent for publication

All authors agreed to submission of the manuscript

Competing interests

AB is a member of the Editorial board of BMC Genomics and a Guest Editor of the collection "Omics of parasites 2025".

Author details

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

²Czech Academy of Sciences, Institute of Parasitology, České Budějovice 370 05, Czech Republic

³Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava 710 00, Czech Republic

⁴Faculty of Science, University of South Bohemia, České Budějovice 370 05, Czech Republic

Received: 8 February 2025 / Accepted: 2 September 2025

Published online: 01 December 2025

References

- Maslov DA, Votýpka J, Yurchenko V, Lukeš J. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. Trends Parasitol. 2013;29:43–52. <https://doi.org/10.1016/j.pt.2012.11.001>.
- Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. Evolution of parasitism in kinetoplastid flagellates. Mol Biochem Parasitol. 2014;195:115–22. <https://doi.org/10.1016/j.molbiopara.2014.05.007>.
- Lukeš J, Butenko A, Hashimi H, Maslov DA, Votýpka J, Yurchenko V. Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. Trends Parasitol. 2018;34(6):466–80. <https://doi.org/10.1016/j.pt.2018.03.002>.
- McGwire BS, Satoskar AR. Leishmaniasis: clinical syndromes and treatment. QJM Int J Med. 2014;107:7–14. <https://doi.org/10.1093/qjmed/hct116>.

5. Steverding D. The history of leishmaniasis. *Parasit Vectors*. 2017;10:82. <https://doi.org/10.1186/s13071-017-2028-5>.
6. Steverding D. The history of African trypanosomiasis. *Parasit Vectors*. 2008;1:3. <https://doi.org/10.1186/1756-3305-1-3>.
7. Kraeva N, Butenko A, Hlaváčová J, Kostygov A, Myškova J, Grybchuk D, et al. *Leptomonas seymouri*: adaptations to the dixenous life cycle analyzed by genome sequencing, transcriptome profiling and co-infection with *Leishmania donovani*. *PLoS Pathog*. 2015;11:e1005127. <https://doi.org/10.1371/journal.ppat.1005127>.
8. Kostygov AY, Albanaz ATS, Butenko A, Gerasimov ES, Lukeš J, Yurchenko V. Phylogenetic framework to explore trait evolution in Trypanosomatidae. *Trends Parasitol*. 2024;40:96–9. <https://doi.org/10.1016/j.pt.2023.11.009>.
9. Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, Field MC, et al. Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep*. 2016;6:23704. <https://doi.org/10.1038/srep23704>.
10. Jaskowska E, Butler C, Preston G, Kelly S. *Phytomonas*. Trypanosomatids adapted to plant environments. *PLoS Pathog*. 2015;11:e1004484. <https://doi.org/10.1371/journal.ppat.1004484>.
11. Clayton C. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol*. 2019;9:190072. <https://doi.org/10.1098/rsob.190072>.
12. Maslov DA, Oppenheer FR, Kostygov AY, Hashimi H, Lukeš J, Yurchenko V. Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology*. 2019;146:1–27. <https://doi.org/10.1017/S0031182018000951>.
13. Albanaz ATS, Gerasimov ES, Shaw JJ, Sádlová J, Lukeš J, Volf P, et al. Genome analysis of *Endotrypanum* and *Porcisia* spp., closest phylogenetic relatives of *Leishmania*, highlights the role of amastins in shaping pathogenicity. *Genes*. 2021;12(3):444. <https://doi.org/10.3390/genes12030444>.
14. Reis-Cunha JL, Pimenta-Carvalho SA, Almeida LV, Coqueiro-dos-Santos A, Marques CA, Black JA, et al. Ancestral aneuploidy and stable chromosomal duplication resulting in differential genome structure and gene expression control in trypanosomatid parasites. *Genome Res*. 2024;34:441–53. <https://doi.org/10.1101/gr.278550.123>.
15. Rastrojo A, García-Hernández R, Vargas P, Camacho E, Corvo L, Imamura H, et al. Genomic and transcriptomic alterations in *Leishmania donovani* lines experimentally resistant to antileishmanial drugs. *Int J Parasitol Drugs Drug Resist*. 2018;8:246–64. <https://doi.org/10.1016/j.ijpddr.2018.04.002>.
16. Fiebig M, Kelly S, Gluenz E. Comparative life cycle transcriptomics revises *Leishmania mexicana* genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLoS Pathog*. 2015;11:e1005186. <https://doi.org/10.1371/journal.ppat.1005186>.
17. Mannaert A, Downing T, Imamura H, Dujardin J-C. Adaptive mechanisms in pathogens: universal aneuploidy in *Leishmania*. *Trends Parasitol*. 2012;28:370–6. <https://doi.org/10.1016/j.pt.2012.06.003>.
18. Schwarz RS, Bauman GR, Murphy CA, Ravoet J, de Graaf DC, Evans JD. Characterization of two species of Trypanosomatidae from the honey bee *Apis mellifera*: *Crithidia mellifica* Langridge and mcghee, and *Lotmaria passim* n. gen., n. sp. *J Eukaryot Microbiol*. 2015;62:567–83. <https://doi.org/10.1111/jeu.12209>.
19. Sadd BM, Barribeau SM. Heterogeneity in infection outcome: lessons from a bumblebee-trypanosome system. *Parasite Immunol*. 2013;35:339–49. <https://doi.org/10.1111/pim.12043>.
20. Brown MJF, Schmid-Hempel R, Schmid-Hempel P. Strong context-dependent virulence in a host–parasite system: reconciling genetic evidence with theory. *J Anim Ecol*. 2003;72:994–1002. <https://doi.org/10.1046/j.1365-2656.2003.00770.x>.
21. Brown MJF, Loosli R, Schmid-Hempel P. Condition-dependent expression of virulence in a trypanosome infecting bumblebees. *Oikos*. 2000;91:421–7. <https://doi.org/10.1034/j.1600-0706.2000.910302.x>.
22. Buendía-Abad M, García-Palencia P, de Pablos LM, Alunda JM, Osuna A, Martín-Hernández R, et al. First description of *Lotmaria passim* and *Crithidia mellifica* haptomonad stages in the honeybee hindgut. *Int J Parasitol*. 2022;52:65–75. <https://doi.org/10.1016/j.ijpara.2021.06.005>.
23. Carreira de Paula J, García Olmedo P, Gómez-Moracho T, Buendía-Abad M, Higes M, Martín-Hernández R, et al. Promastigote EPS secretion and haptomonad biofilm formation as evolutionary adaptations of trypanosomatid parasites for colonizing honeybee hosts. *Npj Biofilms Microbiomes*. 2024;10:1–11. <https://doi.org/10.1038/s41522-024-00492-x>.
24. Arismendi N, Castro MP, Vargas M, Zapata C, Riveros G. The trypanosome *Lotmaria passim* prevails in honey bees of different ages and stages of development. *J Apic Res*. 2020;7. <https://doi.org/10.1080/00218839.2020.1828239>.
25. Ravoet J, Maharramov J, Meeus I, De Smet L, Wenseleers T, Smagghe G, et al. Comprehensive bee pathogen screening in Belgium reveals *Crithidia mellifica* as a new contributory factor to winter mortality. *PLoS ONE*. 2013;8:e72443.
26. Stevanovic J, Schwarz RS, Vojnovic B, Evans JD, Irwin RE, Glavinic U, et al. Species-specific diagnostics of *Apis mellifera* trypanosomatids: a nine-year survey (2007–2015) for trypanosomatids and microsporidians in Serbian honey bees. *J Invertebr Pathol*. 2016;139:6–11. <https://doi.org/10.1016/j.jip.2016.07.001>.
27. Morimoto T, Kojima Y, Yoshiyama M, Kimura K, Yang B, Peng G, et al. Molecular detection of protozoan parasites infecting *Apis mellifera* colonies in Japan. *Environ Microbiol Rep*. 2013;5:74–7. <https://doi.org/10.1111/j.1758-2229.2012.00385.x>.
28. Bartolomé C, Buendía-Abad M, Benito M, Sobrino B, Amigo J, Carracedo A, et al. Longitudinal analysis on parasite diversity in honeybee colonies: new taxa, high frequency of mixed infections and seasonal patterns of variation. *Sci Rep*. 2020;10(1):10454. <https://doi.org/10.1038/s41598-020-67183-3>.
29. Hall RJ, Prager H, Phiri BJ, Fan Q-H, Li X, Parnell A, et al. Apicultural practice and disease prevalence in *Apis mellifera*, New Zealand: a longitudinal study. *J Apic Res*. 2021;60:644–58. <https://doi.org/10.1080/00218839.2021.1936422>.
30. Runckel C, Flenniken ML, Engel JC, Ruby JG, Ganem D, Andino R, et al. Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. *PLoS ONE*. 2011;6:e20656. <https://doi.org/10.1371/journal.pone.0020656>.
31. Cornman RS, Tarpy DR, Chen Y, Jeffreys L, Lopez D, Pettis JS, et al. Pathogen webs in collapsing honey bee colonies. *PLoS ONE*. 2012;7:e43562. <https://doi.org/10.1371/journal.pone.0043562>.
32. Liu Q, Lei J, Darby AC, Kadowaki T. Trypanosomatid parasite dynamically changes the transcriptome during infection and modifies honey bee physiology. *Commun Biol*. 2020;3:1–8. <https://doi.org/10.1038/s42003-020-0775-x>.
33. Gómez-Moracho T, Buendía-Abad M, Benito M, García-Palencia P, Barrios L, Bartolomé C, et al. Experimental evidence of harmful effects of *Crithidia mellifica* and *Lotmaria passim* on honey bees. *Int J Parasitol*. 2020;50:1117–24. <https://doi.org/10.1016/j.ijpara.2020.06.009>.
34. Alcolea PJ, Alonso A, Gómez MJ, Sánchez-Gorostiaga A, Moreno-Paz M, González-Pastor E, et al. Temperature increase prevails over acidification in gene expression modulation of amastigote differentiation in *Leishmania infantum*. *BMC Genomics*. 2010;11:31. <https://doi.org/10.1186/1471-2164-11-31>.
35. Fahrenholz L, Lamprecht I, Schricker B. Thermal investigations of a honey bee colony: thermoregulation of the hive during summer and winter and heat production of members of different bee castes. *J Comp Physiol B*. 1989;159:551–60. <https://doi.org/10.1007/BF00694379>.
36. Esch H. Über die Körpertemperaturen und den Wärmehaushalt von *Apis mellifica*. *Z Für Vgl Physiol*. 1960;43:305–35. <https://doi.org/10.1007/BF00298066>.
37. Zheng H, Powell JE, Steele MI, Dietrich C, Moran NA. Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proc Natl Acad Sci U S A*. 2017;114:4775–80. <https://doi.org/10.1073/pnas.1701819114>.
38. Zilberstein D, Shapira M. The role of pH, and temperature in the development of leishmania parasites. *Annu Rev Microbiol*. 1994;48:449–70. <https://doi.org/10.1146/annurev.mi.48.100194.002313>.
39. Pan AA, Dubois SM, Eperon S, Rivas L, Hodgkinson V, Traub-Cseko Y, et al. Developmental life cycle of *Leishmania*—Cultivation and characterization of cultured extracellular amastigotes. *J Eukaryot Microbiol*. 1993;40:213–23. <https://doi.org/10.1111/j.1550-7408.1993.tb04906.x>.
40. Markowitz LM, Nearman A, Zhao Z, Boncristiani D, Butenko A, de Pablos LM, et al. Somy evolution in the honey bee infecting trypanosomatid parasite, *Lotmaria passim*. *G3: Genes, Genomes, Genetics*. 2024;jkae258. <https://doi.org/10.1093/g3journal/jkae258>.
41. Runckel C, DeRisi J, Flenniken ML. A draft genome of the honey bee trypanosomatid parasite *Crithidia mellifica*. *PLoS ONE*. 2014;9:e95057. <https://doi.org/10.1371/journal.pone.0095057>.
42. Davey JW, Catta-Preta CMC, James S, Forrester S, Motta MCM, Ashton PD, et al. Chromosomal assembly of the nuclear genome of the endosymbiont-bearing trypanosomatid *Angomonas deanei*. *G3 Genes Genomes Genet*. 2020;11:jkaa018. <https://doi.org/10.1093/g3journal/jkaa018>.
43. Soderlund C, Bomhoff M, Nelson WM. SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res*. 2011;39:e68. <https://doi.org/10.1093/nar/gkr123>.

44. Shen G, Wang W-L. Cirlize package in R and analytic hierarchy process (AHP): contribution values of ABCDE and AGL6 genes in the context of floral organ development. *PLoS ONE*. 2022;17:e0261232. <https://doi.org/10.1371/journal.pone.0261232>.
45. Brown MR, Manuel G, de La Rosa P, Blaxter M. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics*. 2025;41:btaf049. <https://doi.org/10.1093/bioinformatics/btaf049>.
46. Poláková E, Záhonová K, Albanaz ATS, Butenko A, Lukeš J, Yurchenko V. Diverse telomeres in trypanosomatids. *Parasitology*. 2021;148:1254–70. <https://doi.org/10.1017/S0031182021000378>.
47. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
48. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 2017;8:28–36. <https://doi.org/10.1111/2041-210X.12628>.
49. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R Language. *Bioinformatics*. 2004;20:289–90. <https://doi.org/10.1093/bioinformatics/btg412>.
50. R Core Team. R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
51. Brooks ME, Kristensen K, van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. GlimmTMB balances speed and flexibility among packages for Zero-inflated generalized linear mixed modeling. *R J*. 2017;9:378–400.
52. Lenth RV. Least-squares means: the R package lsmeans. *J Stat Softw*. 2016;69:1–33. <https://doi.org/10.18637/jss.v069.i01>.
53. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009.
54. Salathé R, Tognazzo M, Schmid-Hempel R, Schmid-Hempel P. Probing mixed-genotype infections I: extraction and cloning of infections from hosts of the trypanosomatid *Crithidia bombi*. *PLoS ONE*. 2012;7:e49046. <https://doi.org/10.1371/journal.pone.0049046>.
55. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
56. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
57. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
58. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
59. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25:25–9. <https://doi.org/10.1038/75556>.
60. The Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The gene ontology knowledgebase in 2023. *Genetics*. 2023;224:iyad031. <https://doi.org/10.1093/genetics/iyad031>.
61. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–62. <https://doi.org/10.1093/nar/gkv1070>.
62. Törönen P, Holm L. PANNZER—a practical tool for protein function prediction. *Protein Sci*. 2022;31(1):118–28. <https://doi.org/10.1002/pro.4193>.
63. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. ClusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov (Camb)*. 2021;2:100141. <https://doi.org/10.1016/j.xinn.2021.100141>.
64. Bianca F, Ispano E, Gazzola E, Lavezzo E, Fontana P, Toppo S. FunTaxIS-lite: a simple and light solution to investigate protein functions in all living organisms. *Bioinformatics*. 2023;39:btad549. <https://doi.org/10.1093/bioinformatics/btad549>.
65. Kanehisa M, Sato Y, Morishima K. Blastkoala and ghostkoala: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428:726–31. <https://doi.org/10.1016/j.jmb.2015.11.006>.
66. Camacho E, González-de la Fuente S, Solana JC, Rastrojo A, Carrasco-Ramiro F, Requena JM, et al. Gene annotation and transcriptome delineation on a *de novo* genome assembly for the reference *Leishmania major* Friedlin strain. *Genes*. 2021;12:1359. <https://doi.org/10.3390/genes12091359>.
67. Assis LHC, Andrade-Silva D, Shiburah ME, de Oliveira BCD, Paiva SC, Abuchery BE, et al. Cell cycle, telomeres, and telomerase in *Leishmania* spp. what do we know so far? *Cells*. 2021;10:103195. <https://doi.org/10.3390/cells10113195>.
68. Fu G, Barker DC. Characterisation of *Leishmania* telomeres reveals unusual telomeric repeats and conserved telomere-associated sequence. *Nucleic Acids Res*. 1998;26:2161–7.
69. Martínez-Calvillo S, Romero-Meza G, Vizuet-de-Rueda JC, Florencio-Martínez LE, Manning-Cela R, Nepomuceno-Mejía T. Epigenetic regulation of transcription in trypanosomatid protozoa. *Curr Genomics*. 2018;19:140–9. <https://doi.org/10.2174/1389202918666170911163517>.
70. Bringaud F, Rivière L, Coustou V. Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol Biochem Parasitol*. 2006;149:1–9. <https://doi.org/10.1016/j.molbiopara.2006.03.017>.
71. Vannette RL, Mohamed A, Johnson BR. Forager bees (*Apis mellifera*) highly express immune and detoxification genes in tissues associated with nectar processing. *Sci Rep*. 2015;5:16224. <https://doi.org/10.1038/srep16224>.
72. Povelones ML, Holmes NA, Povelones M. A sticky situation: when trypanosomatids attach to insect tissues. *PLoS Pathog*. 2023;19:e1011854. <https://doi.org/10.1371/journal.ppat.1011854>.
73. Filosa JN, Berry CT, Ruthel G, Beverley SM, Warren WC, Tomlinson C, et al. Dramatic changes in gene expression in different forms of *Crithidia fasciculata* reveal potential mechanisms for insect-specific adhesion in kinetoplastid parasites. *PLoS Negl Trop Dis*. 2019;13:e0007570. <https://doi.org/10.1371/journal.pntd.0007570>.
74. Sloan MA, Brooks K, Otto TD, Sanders MJ, Cotton JA, Ligoxygakis P. Transcriptional and genomic parallels between the monoxenous parasite *Herpetomonas muscarum* and *Leishmania*. *PLoS Genet*. 2019;15:e1008452. <https://doi.org/10.1371/journal.pgen.1008452>.
75. Saxena A, Lahav T, Holland N, Aggarwal G, Anupama A, Huang Y, et al. Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Mol Biochem Parasitol*. 2007;152:53–65. <https://doi.org/10.1016/j.molbio.2006.11.011>.
76. Ouellette M, Légaré D, Haimeur A, Grondin K, Roy G, Brochu C, et al. ABC transporters in *Leishmania* and their role in drug resistance. *Drug Resist Updat*. 1998;1:43–8. [https://doi.org/10.1016/S1368-7646\(98\)80213-6](https://doi.org/10.1016/S1368-7646(98)80213-6).
77. Palmer-Young EC, Farrell IW, Adler LS, Milano NJ, Egan PA, Junker RR, et al. Chemistry of floral rewards: intra- and interspecific variability of nectar and pollen secondary metabolites across taxa. *Ecol Monogr*. 2019;89:e01335. <https://doi.org/10.1002/ecm.1335>.
78. Palmer-Young EC, Schwarz RS, Chen Y, Evans JD. Can floral nectars reduce transmission of *Leishmania*? *PLoS Negl Trop Dis*. 2022;16:e0010373. <https://doi.org/10.1371/journal.pntd.0010373>.
79. Johnson RM, Harpur BA, Dogantzis KA, Zayed A, Berenbaum MR. Genomic footprint of evolution of eusociality in bees: floral food use and cypome blooms. *Insectes Soc*. 2018;65:445–54. <https://doi.org/10.1007/s00040-018-0631-x>.
80. Ennes-Vidal V, Menna-Barreto RFS, Santos ALS, Branquinho MH, d'Avila-Levy CM. MDL28170, a calpain inhibitor, affects *Trypanosoma cruzi* metacyclogenesis, ultrastructure and attachment to *Rhodnius prolixus* midgut. *PLoS ONE*. 2011;6:e18371. <https://doi.org/10.1371/journal.pone.0018371>.
81. Cunningham AC. Parasitic adaptive mechanisms in infection by *Leishmania*. *Exp Mol Pathol*. 2002;72:132–41. <https://doi.org/10.1006/exmp.2002.2418>.
82. Mottram JC, Coombs GH, Alexander J. Cysteine peptidases as virulence factors of *Leishmania*. *Curr Opin Microbiol*. 2004;7:375–81. <https://doi.org/10.1016/j.mib.2004.06.010>.
83. Wang L, Sloan MA, Ligoxygakis P. Intestinal NF- κ B and STAT signalling is important for uptake and clearance in a *Drosophila*-*Herpetomonas* interaction model. *PLoS Genet*. 2019;15:e1007931. <https://doi.org/10.1371/journal.pgen.1007931>.
84. Schwarz RS, Evans JD. Single and mixed-species trypanosome and microsporidia infections elicit distinct, ephemeral cellular and humoral immune responses in honey bees. *Dev Comp Immunol*. 2013;40:300–10. <https://doi.org/10.1016/j.dci.2013.03.010>.
85. Brunner FS, Schmid-Hempel P, Barribeau SM. Immune gene expression in *Bombus terrestris*: signatures of infection despite strong variation among populations, colonies, and sister workers. *PLoS ONE*. 2013;8:e68181. <https://doi.org/10.1371/journal.pone.0068181>.
86. Barribeau SM, Sadd BM, du Plessis L, Schmid-Hempel P. Gene expression differences underlying genotype-by-genotype specificity in a host–parasite system. *Proc Natl Acad Sci U S A*. 2014;111:3496–501. <https://doi.org/10.1073/pnas.1318628111>.

87. Lahav T, Sivam D, Volpin H, Ronen M, Tsigankov P, Green A, et al. Multiple levels of gene regulation mediate differentiation of the intracellular pathogen *Leishmania*. *FASEB J*. 2011;25(2):515–25. <https://doi.org/10.1096/fj.10-157529>.
88. Yuan X, Sun J, Kadowaki T. Aspartyl protease in the secretome of honey bee trypanosomatid parasite contributes to infection of bees. *Parasit Vectors*. 2024;17:60. <https://doi.org/10.1186/s13071-024-06126-7>.
89. Schmid-Hempel P, Aebi M, Barribeau S, Kitajima T, Plessis L, du, Schmid-Hempel R, et al. The genomes of *Crithidia bombi* and *C. expoeki*, common parasites of bumblebees. *PLoS ONE*. 2018;13:e0189738. <https://doi.org/10.1371/journal.pone.0189738>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.