

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

José Franklin Calderón Tantaleán

**Análise genômica comparativa e evolutiva do genoma do monoxeno *Blechnomonas* sp.
TCC303E**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

São Paulo

2022

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS BIOMÉDICAS
PROGRAMA DE PÓS-GRADUAÇÃO INTERUNIDADES EM
BIOINFORMÁTICA

Análise genômica comparativa e evolutiva do genoma do monoxeno
***Blechromonas* sp. TCC303E**

Autor: José Franklin Calderón Tantaleán– Bacharel em Ciências Biológicas (Universidad Nacional Pedro Ruiz Gallo)

Orientador: Prof. Dr. João Marcelo Pereira Alves

Programa: Interunidades em bioinformática

E-mail: jcalderonta@usp.br

Área do conhecimento: Ciências Biológicas (Genômica, Bioinformática)

Departamento de Parasitologia, ICB, Universidade de São Paulo.

E-mail: jotajj@usp.br

São Paulo
2022

Nome: José Franklin Calderón Tantaleán

Título: Análise genômica comparativa e evolutiva do genoma do monoxeno *Blechromonas* sp.
TCC303E

Dissertação apresentada ao Programa Interunidades em
Bioinformática da Universidade de São Paulo para obtenção do título
de Mestre em Ciências.

Aprovado em:

Banca Examinadora

Prof.(a) Dr.(a) _____

Instituição _____

Julgamento _____

Prof.(a) Dr.(a) _____

Instituição _____

Julgamento _____

Prof.(a) Dr.(a) _____

Instituição _____

Julgamento _____

À toda minha família, em especial minha mãe

AGRADECIMENTOS

Acima de tudo, agradeço ao meu Deus, por seu mistério infinito onde dia a dia nós cientistas trabalhamos para entendê-lo um pouco mais.

A minha mãe Lila Tantalean Campos e minha irmã Leidy Yannina Calderon Tantalean por serem minha força e empolgação para ser melhor a cada dia e a esse amor infinito que nos une, e também agradeço a meu amor de pandemia (agora noiva) Samara Spagnolo Sousa por sua parceria e companhia que me faz sentir bem cada dia. Agradeço também ao meu melhor amigo Hachi, por todos os momentos de diversão e tristeza juntos e aquele amor incondicional que os cães têm por seus donos, descanse em paz meu amigo.

Ao Prof. João Marcelo Pereira Alves, mais conhecido pelo “J” por abrir as portas de seu laboratório para mim, pela oportunidade de fazer parte do time Labo 40; pela paciência e a confiança depositada para o desenvolvimento deste trabalho.

Ao pessoal do lab. 40, Percy Omar Túllume Vergara e Anderson Carvalho Morais, obrigado pelo apoio e das dicas neste trabalho de mestre e também pela sua amizade. A Guillermo Uceda Campos pelas conversas e apoio sobre diversos assuntos.

À Universidade de São Paulo, às agências de fomento à pesquisa CAPES pelo apoio financeiro.

E para você, que embora seu nome não esteja listado nestes agradecimentos, mais que também contribui de alguma forma ou outra no desenvolvimento deste projeto.

Só tenho a agradecer a todos vocês, muito obrigado!

RESUMO

O conhecimento dos tripanossomatídeos vem principalmente dos heteroxenos que foram muito mais estudados por terem impactos negativos na sociedade, deixando o maior grupo de tripanossomatídeos (monoxenos) quase esquecido por muito tempo. No entanto, os monoxenos se tornaram relevantes recentemente por serem bons modelos para entender a origem do parasitismo e a evolução dos heteroxenos, devido a serem basais a todos os heteroxenos importantes e compartilharem características genéticas comuns, além de co-infectar os hospedeiros desses heteroxenos. Diante dessa necessidade, este estudo tem como objetivo caracterizar o genoma do monoxeno *Blechromonas* sp. TCC303E (Classe Kinetoplastea; subfamília Blechomonadinae) através de técnicas bioinformáticas para analisar aspectos genômicos e evolutivos desse organismo, expandindo assim a disponibilidade de dados genéticos deste gênero, que até agora tinha disponível apenas a sequência genômica da espécie-tipo, *Blechromonas ayalai*. *Blechromonas* são parasitas extremamente polimórficos e exclusivos das pulgas (Siphonaptera), as quais podem transmitir agentes etiológicos de doenças veterinárias e médicas. Nossos resultados mostram uma montagem com uma cobertura média de 38,6 vezes, de boa qualidade e composta de 728 *scaffolds*, com N50 de ~61 kbp e uma média de GC de 50,3%. Um dos *scaffolds* contém DNA mitocondrial, com uma média de GC de 19,7%. Foram preditos 8.044 CDS, das quais foram anotadas funcionalmente 7.002, além de genes de RNAs: 71 tRNAs, 4 rRNAs, 62 RNAs não codificantes. A análise de sintenia mostrou que *Blechromonas* mantém uma arquitetura genômica semelhante à dos demais tripanossomatídeos, formada por grandes blocos sintênicos altamente conservados e consistentes com suas distâncias filogenéticas. Identificamos 10.753 grupos ortólogos (GOs) entre os 12 organismos usados neste estudo, com 2803 GOs de cópia única e 6992 proteínas específicas de espécies agrupadas em 1497 GOs. A análise filogenômica empregando os 2803 GOs de cópia única, baseada em modelos de coalescência de múltiplas espécies, reafirma uma origem precoce do gênero *Blechromonas*, como um clado irmão dos tripanossomatídeos, exceto por Trypanosomatinae (*Trypanosoma* spp.) e Paratrypanosomatinae (*Paratrypanosoma* spp.). As distribuições das categorias funcionais e de domínio de *Blechromonas* sp. TCC303E mostraram tendências similares aos outros tripanossomatídeos, com reduções nas redundâncias funcionais e expansões dos caracteres ancestrais, especialmente aquelas famílias multigênicas que possibilitam as mudanças morfológicas, aproveitamento de nutrientes e interação com os hospedeiros. Em conclusão, em *Blechromonas* sp. TCC303E, como nos demais tripanossomatídeos, as distribuições assimétricas dos repertórios genéticos, especialmente aquelas das famílias multigênicas de superfície, peptidases e transposons seriam as adaptações que provavelmente possibilitaram a sobrevivência no seu hospedeiro.

Palavras chaves: *Blechromonas*, monoxenos, genômica, filogenômica.

ABSTRACT

Calderon-Tantalean, J. F. **Comparative and evolutionary analysis of the genome of the monoxenous trypanosomatid *Blechromonas* sp. TCC303E.** Dissertação (Mestrado em Bioinformática) - Programa Interunidades em Bioinformática, Universidade de São Paulo, São Paulo, 2022.

Knowledge about trypanosomatids derives mainly from the heteroxenous species, which have been studied much more due to their negative impacts on society, leaving the largest trypanosomatid group (the monoxenous ones) nearly forgotten for a long time. However, the monoxenous species have recently become relevant due to being good models for understanding the origin of parasitism and the origin of the heteroxenous species. That is a consequence of the monoxenous being basal to the most important heteroxenous species, sharing many genetic features, as well as co-infecting the same hosts. With that in mind, this study aims at characterizing the genome of the monoxenous species *Blechromonas* sp. TCC303E (Class Kinetoplastea; subfamily Blechomonadinae) by bioinformatic techniques, analyzing genomic and evolutionary aspects of this organism, thus expanding the availability of genetic data about this genus, which until now had only one genomic sequence available, that of the type-species, *Blechromonas ayalai*. *Blechromonas* spp. are extremely polymorphic parasites that are restricted to fleas (Siphonaptera), which can transmit the etiologic agents of several medical and veterinary diseases. Our results include a good-quality assembly with average coverage of 38.6 fold and composed of 728 scaffolds, with an N50 of ~61 kbp and average GC composition of 50.3%. One of the scaffolds contains the mitochondrial DNA, with an average GC of 19.7%. Gene prediction yielded 8,044 CDS, of which 7,002 received functional annotation, as well as RNA genes: 71 tRNAs, 4 rRNAs, 62 other non-coding RNAs. Synteny analysis showed that *Blechromonas* conserves a genomic architecture that is similar to that of the other trypanosomatids, characterized by large and highly conserved syntenic blocks, with conservation inversely proportional to phylogenetic distances. We identified 10,753 orthologous groups (OGs) among the 12 organisms analyzed in this study; 2,803 OGs were single-copy and 6,992 species-specific proteins were grouped in 1,497 OGs. Phylogenomic analysis using the 2,803 single-copy OGs, based on multiple species coalescence models, reaffirms an early origin for the genus *Blechromonas*, as a sister clade of all trypanosomatids excluding Trypanosomatinae (*Trypanosoma* spp.) and Paratrypanosomatinae (*Paratrypanosoma* spp.). Functional category and domain distribution in *Blechromonas* sp. TCC303E showed similar trends to other trypanosomatids, with reduction in functional redundancies and expansion of ancestral characters, especially the multigene families that enable morphological changes, acquisition and processing of nutrients, and interaction with the hosts. In conclusion, in *Blechromonas* sp. TCC303E, as in the other trypanosomatids, the asymmetric distributions of genetic repertoires, especially the multigene families for surface proteins, peptidases and transposons, would be the adaptations that are likely to have enabled survival in the host.

Keywords: *Blechromonas*, monoxenous, genomics, phylogenomics.

LISTA DE ILUSTRAÇÕES

- Figura 1.** Comparação das categorias BUSCO para avaliar a completude da montagem de *Blechomonas* sp. TCC303E com respeito aos genomas de *B. ayalai* e *Leishmania major*.....26
- Figura 2.** Gráfico dotplot gerado pelo pacote MUMmerplot para a comparação estrutural usando sequências nucleotídicas da montagem de *Blechomonas* sp. TCC303E com respeito ao genoma de *B. ayalai*.....27
- Figura 3.** Gráfico dotplot gerado pelo pacote MUMmerplot para a comparação estrutural usando sequências de aminoácidos da montagem de *Blechomonas* sp. TCC303E com respeito ao genoma de *B. ayalai*.....28
- Figura 4.** Gráfico da sintenia entre as espécies *Blechomonas* sp. TCC303E, *B. ayalai* e *L. major*.....33
- Figura 5.** Comparação das quantidades dos proteomas de 9 tripanossomatídeos e de *B. saltans* nas categorias funcionais KOG geradas pela ferramenta eggNOG.....34
- Figura 6.** Comparação dos domínios protéicos anotados com domínios PFAM de 11 espécies de tripanossomatídeos e de *B. saltans*.....39
- Figura 7.** Comparação do número de grupos ortólogos compartilhados pelas 11 espécies de tripanossomatídeos e *B. saltans* gerados pela ferramenta OrthoFinder.....45
- Figura 8.** Heatmap dos 23 grupos ortólogos mais abundantes com anotação conhecida e presentes em todas as 11 espécies de tripanossomatídeos e *B. saltans* analisadas com o programa OrthoFinder.....46
- Figura 9.** Árvore de espécies de 2803 grupos ortólogos de cópia única, inferida mediante modelos de coalescência de múltiplas espécies.....51
- Figura 10.** Mapa metabólico que mostra a presença e ausência de genes da rota metabólica de biosíntese de moléculas de porfirina de quatro espécies de tripanossomatídeos.....53
- Figura 11.** Mapa metabólico que mostra a presença e ausência de genes da rota metabólica dos aminoácidos alanina, ácido aspártico e ácido glutâmico de quatro espécies de tripanossomatídeos.....54
- Figura 12.** Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos arginina e prolina de quatro espécies de tripanossomatídeos.....55
- Figura 13.** Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos cisteína e metionina de quatro espécies de tripanossomatídeos.....56
- Figura 14.** Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos glicina, serina e metionina de quatro espécies de tripanossomatídeos.....57

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1. Sequências genômicas das 17 espécies de tripanossomatídeos..... | 21 |
| Tabela 2. Principais estatísticas da montagem do genoma de <i>Blechromonas</i> sp. TCC303E..... | 25 |
| Tabela 3. Genes encontrados no maxicírculo de <i>Blechromonas</i> sp. TCC303E usando a ferramenta BLASTx ... | 26 |
| Tabela 4. Resumo das características do genomas de 10 espécies de tripanossomatídeos..... | 31 |
| Tabela 5. Descrição das categorias dos resultados obtidos com o programa OrthoFinder para 11 espécies de tripanossomatídeos e <i>B. saltans</i> | 46 |

SUMÁRIO

| | |
|--|-----------|
| INTRODUÇÃO | 11 |
| II. Justificativa | 18 |
| III. Objetivos | 19 |
| III.1 Objetivo geral | 19 |
| III.2 Objetivos específicos | 19 |
| IV. Metodologia e perspectivas | 19 |
| IV.1 Sequências e montagem do genoma | 19 |
| IV.2 Avaliação da montagem | 20 |
| IV.3 Predição gênica | 21 |
| IV.4 Anotação | 22 |
| IV.5 Genômica comparativa | 22 |
| IV.6 Filogenômica | 23 |
| V. RESULTADOS E DISCUSSÃO | 23 |
| V.1 Características e avaliação da montagem genômica | 23 |
| V.2 Predição de genes codificadores de proteínas e RNAs. | 29 |
| V.3 Anotação funcional | 31 |
| V.4 Análise de sintenia de <i>Blechromonas</i> sp. TCC303E | 32 |
| V.5 Genômica comparativa de <i>Blechromonas</i> sp. TCC303E baseada na classificação funcional | 33 |
| V.6 Comparação genômica baseada nos domínios PFAM | 39 |
| V.7 Análise das famílias de genes usando a abordagem de clustering | 44 |
| V.8 Expansões de família de genes dentro dos Clusters de Grupos Ortólogos | 46 |
| V.9 Filogenômica | 50 |
| V.10. Vias metabólicas | 51 |
| V.10.1 Via biossintética do grupo heme | 52 |
| V.10.2 Aminoácidos | 53 |
| V.12 Conclusão | 58 |
| REFERÊNCIAS | 59 |
| ANEXO | 74 |

INTRODUÇÃO

A classe Kinetoplastea é formada por protozoários que apresentam como sinapomorfia o cinetoplasto, que é uma rede de moléculas de DNA circular concatenadas (os mini- e maxicírculos) dentro da única mitocôndria e está normalmente associado ao corpo basal do flagelo. Além disso, apresentam outras características peculiares da classe que incluem: (i) um complexo sistema de edição de RNA mitocondrial, com gasto de energia; (ii) trans-splicing em todos os transcritos de mRNA; (iii) genes agrupados em grupos policistrônicos; (iv) modificações únicas de nucleotídeos; (v) presença de glicossomas (compartimento celular com enzimas peroxissomais e para a glicólise) (Simpson et al. 2006). Nesta classe encontramos uma grande abundância e diversidade de organismos, que só são superadas pelo filo Apicomplexa, e em gama de hospedeiros, é superada apenas por nematóides parasitas (Pawlowski et al. 2012; Vickerman 1994). Isto lhes permitiu serem distribuídos amplamente e muito adaptáveis. Nesta classe, como nas principais linhagens eucarióticas (Apicomplexa, Microsporidia e Neodermata), o parasitismo também está presente, e surgiu de forma independente várias vezes com uma distribuição não uniforme em suas cinco ordens: de modo intermitente, como nas ordens Eubodonida, Parabodonida, Neobodonida (organismos principalmente de vida livre, chamados bodonídeos), ou de forma constante, como nas ordens Prokinetoplastida e Trypanosomatida, que abrigam apenas organismos parasitas (Lukeš et al. 2014; Maslov et al. 2001; Simpson et al. 2006). Desta forma, a distribuição dos estilos de vida também é uma característica interessante desta classe, se o que se pretende é entender a origem do parasitismo, uma vez que cada grupo de parasitas possui parentes de vida livre que são filogeneticamente próximos, então a classe Kinetoplastea abriga um bom cenário para entender a origem do parasitismo.

A família Trypanosomatidae (ordem: Trypanosomatida), é uma grande família de parasitas uniflagelados obrigatórios que concentra a maior diversidade de cinetoplastídeos e número de hospedeiros, infectando invertebrados e todas as classes de vertebrados e plantas (Jackson 2015; Stevens et al. 1999). Esses parasitas têm dois tipos de ciclos de vida: monoxênicos, que requerem somente um hospedeiro para completar seu ciclo biológico, ou heteroxênicos, que também requerem um organismo intermediário (geralmente insetos) para infectar seu hospedeiro, que pode ser uma planta ou vertebrado (Maslov et al. 2001). Ambos os ciclos de vida podem ser complexos (especialmente em heteroxenos), que manifestam diversos fenótipos e morfotipos; sendo os principais morfotipos: tripomastigota, promastigota, amastigota, opistomastigota, epimastigota e coanomastigota (Alcolea et al. 2014; Wheeler et al. 2013).

Além do tipo de ciclo de vida e morfotipo, a taxonomia atual dos tripanossomatídeos usa um sistema de “códigos de barras” moleculares, complementado por outras informações taxonômicas relevantes, como a identidade do hospedeiro e a localização da espécie dentro do hospedeiro. Para a

família Trypanosomatidae, o sistema de código de barras molecular usa majoritariamente os genes do SSU rRNA e da gliceraldeído-3-fosfato desidrogenase glicossomal (gGAPDH) para a classificação em níveis superiores, inclusive até gênero, e as sequências do gene spliced leader (SL) de RNA e sua respectiva região intergênica para o nível de espécie. Desta forma, as novas espécies são representadas por meio de proxies ou unidades de tipagem (TU, *Typing units*) (Votýpka et al. 2013). Para distinguir essas TUs individualmente, um limiar de corte de 90% de similaridade entre as sequências foi estabelecido, com base na análise de diferentes *Leishmania* (Maslov et al. 2007; Votýpka et al. 2012a). Por outro lado, a disposição dos nichos biológicos é um fator importante na história evolutiva dos tripanossomatídeos que tem levado a várias vias de infecção e, em muitos casos, baixa especificidade para seus hospedeiros (promíscuos); isso resultou na mesma espécie preferindo diferentes locais de colonização e hospedeiros (Kaufer et al. 2017; Maslov et al. 2013; Votýpka et al. 2012b); Por exemplo, monoxenos capazes de infectar insetos da ordem Hemiptera e Diptera foram relatados, ao contrário de algumas espécies de heteroxênicos, como *Leishmania*, que infectam espécies específicas de mosquito-palha (flebotomíneos) (Votýpka et al. 2013). Outro aspecto a se considerar na descrição de uma espécie é a patogenicidade em seus insetos hospedeiros, sendo considerada sub patogênica, intermediária e patogênica; onde o primeiro tem efeitos adversos em condições desfavoráveis e o patogênico tem efeitos adversos mesmo em condições normais (Schaub 1994).

Os heteroxenos são transmitidos por insetos fitófagos ou hematófagos, principalmente das ordens Diptera e Hemiptera. São os tripanossomatídeos mais bem estudados devido à sua importância na medicina, medicina veterinária e economia, o que permitiu ampliar o conhecimento inicial sobre os tripanossomatídeos. É sabido que a maioria das doenças causadas por tripanossomatídeos é zoonótica, sendo os animais hospedeiros chaves para manter a endemicidade, o que representa um problema de saúde pública em todos os continentes (Kaufer et al. 2017; Lukeš et al. 2018; Sylvia e Aldo 2019). Os heteroxenos possuem uma grande variedade de hospedeiros, incluindo o homem e animais domésticos, nos quais causam doenças importantes por meio de espécies do gênero *Trypanosoma* e *Leishmania*, ou podem causar sérios danos na agricultura por meio de espécies do gênero *Phytomonas*, que afetam as plantas (Kaufer et al. 2017; Lukeš et al. 2018; Sylvia e Aldo 2019); (Rangel et al. 2019). Esses organismos diferem muito em suas estratégias de sobrevivência e ciclos de vida. Por exemplo, as espécies icônicas *Trypanosoma cruzi* (doença de Chagas), *Trypanosoma brucei* (doença do sono) e *Leishmania* spp (leishmaniose), também conhecidos como Tritryp (Berriman et al. 2005; El-Sayed et al. 2005a; Ivens et al. 2005), têm ciclos de vida zoonóticos ou antroponóticos e são transmitidos por insetos hematófagos. *T. brucei* passa por um complexo desenvolvimento para produzir formas flageladas infecciosas nas glândulas salivares da mosca tsé-tsé, para posteriormente se espalhar na sangue do hospedeiro vertebrado com variações antigênicas para evitar a resposta imune do hospedeiro. Por sua vez, o *T. cruzi* é transmitido pelas fezes de um triatomíneo e estabelece uma infecção crônica em seu hospedeiro por meio da disseminação intracelular. Já as leishmanias são

transmitidas a seus hospedeiros mamíferos e répteis pelo mosquito-palha, onde evitam a resposta imune por meio da propagação dentro dos macrófagos (Kaufer et al. 2017; Lukeš et al. 2018).

O evidente impacto negativo dos heteroxenos na sociedade justificou o estudo intensivo desses organismos, o que deixou os monoxenos quase esquecidos por muito tempo, apesar de representarem a maioria da família Trypanosomatidae. Nesse sentido, pouco se sabe sobre a diversidade, distribuição e biologia desses organismos. Na verdade, a biodiversidade de monoxênicos pode ser maior do que poderia ser estimado com base na amostragem realizada até o momento de uma pequena fração de hospedeiros potenciais de tripanossomatídeos em alguns países (2.500 insetos amostrados de um milhão estimado), que é suportado pelo número crescente de novas espécies descritas, que até o momento são 300 TUs, configurando atualmente 17 gêneros: *Herpetomonas*, *Blastocrithidia*, *Crithidia*, *Leptomonas*, *Wallacemonas*, *Sergeia*, *Strigomonas*, *Angomonas*, *Rhynchoidomonas*, *Paratrypanosoma*, *Kentomonas*, *Lotmaria*, *Lafontella*, *Jaenimonas*, *Zelonia*, *Novymonas*, e *Blechomonas* (Kaufer et al. 2017, 2019; Lukeš et al. 2018; Rangel et al. 2019).

De forma similar, o conhecimento do ciclo de vida de organismos monoxênicos é escasso, sendo uma generalização baseada na descrição do desenvolvimento de poucas espécies, como *Leptomonas ctenocephali* em pulgas e *Strigomonas oncopelti* em percevejos da espécie *Oncopeltus fasciatus* (Kaufer et al. 2017). Os monoxenos parasitam principalmente insetos das ordens Heteroptera e Diptera, e ocasionalmente infectam as ordens Lepidoptera e Siphonaptera; nesses insetos, os monoxenos colonizam todo o trato digestivo, túbulo de Malpighi, glândulas salivares e alguns invadem a hemocele, conforme relatado principalmente em triatomíneos (Lukeš et al. 2018; McGhee e Cosgrove 1980; Votýpka et al. 2013). Ao contrário dos heteroxenos, os monoxenos são em sua maioria não patogênicos ou pouco patogênicos, com algumas exceções, destacando a espécie *Crithidia bombi*, que parasita abelhas da espécie *Bombus terrestris*, usada na polinização (Brown et al. 2003), e *Blastocrithidia triatoma*, que parasita vetores triatomíneos (Hemiptera, Reduviidae) da doença de Chagas (Schaub 1994, 2009); essas espécies estabelecem infecções crônicas e debilitantes em seus hospedeiros, além de gerar mudanças em seu comportamento, consumo de alimentos, atraso no desenvolvimento, resposta imune, capacidade de inanição e redução do tempo de vida e da taxa de reprodução (Lukeš et al. 2018; McGhee e Cosgrove 1980; Votýpka et al. 2013).

A transmissão dos monoxenos é realizada por diversas vias, como coprofagia direta, predação, canibalismo e transmissão vertical pela superfície dos ovos (Kaufer et al. 2017; Maslov et al. 2013; Votýpka et al. 2012b). Além disso, sua transmissão é potencializada devido à promiscuidade exibida por alguns tripanossomatídeos, o que lhes confere grande potencial infectante, sendo uma preocupação em espécies como *C. bombi* que podem infectar outras espécies de abelhas polinizadoras (Colla et al. 2006; Plischuk e Lange 2009). Talvez esse potencial infeccioso tenha possibilitado a adaptabilidade inesperada a novos organismos, demonstrada por algumas espécies de monoxenos

vistos infectando vários organismos, desde plantas até vertebrados (como roedores e animais domésticos), incluindo pacientes com leishmaniasis ou imunocomprometidos (Kraeva et al. 2015; McGhee e Cosgrove 1980; Morio et al. 2008; Pacheco et al. 1998). Essas transmissões incomuns podem nos dar novas perspectivas para entender a origem do ciclo de vida heteroxênico (Sloan et al. 2019), o que é corroborado por vários estudos que demonstraram que todos os heteroxenos importantes são derivados de monoxenos ancestrais, com os quais também compartilham características genéticas comuns (Barratt et al. 2017). Nesse contexto, os monoxênicos se apresentam como bons modelos para o estudo da origem e evolução do parasitismo heteroxênico devido às características interessantes que apresentam, como grande diversidade biológica, distribuição cosmopolita, potencial impacto na fisiologia de seus hospedeiros, e estreita relação evolutiva e genética com heteroxenos (Kaufer et al. 2017; Lukeš et al. 2018; Sylvania e Aldo 2019); (Kaufer et al. 2017; Lukeš et al. 2018; Sylvania e Aldo 2019).

O sistema taxonômico dos tripanossomatídeos teve início há cerca de 60 anos, com base na morfologia, ciclo de vida e tipo de hospedeiro (taxonomia clássica) que permitiu traçar os primeiros esquemas da filogenética dos tripanossomatídeos (Hoare e Wallace 1966; Wallace 1966). Naquela época, a subordem Trypanosomatina foi definida (um flagelo e cinetoplasto reduzido, atual família Trypanosomatidae) e os monoxenos foram considerados “tripanossomatídeos inferiores”, com a suposição de que todos os tripanossomatídeos evoluíram de um monoxênico ancestral comum. No entanto, esse sistema gerou táxons polifiléticos, devido ao fato de alguns táxons, principalmente monoxênicos, não apresentarem características morfológicas peculiares que os definam em nível de espécie ou mesmo de gênero (Jirků et al. 2012). Em seguida, a microscopia eletrônica contribuiu com a taxonomia, mas sem resultados destacados (Souza 2008).

Com o advento da filogenética molecular, a taxonomia atual afirmou e redefiniu alguns táxons antigos, bem como definiu novos, incluindo níveis taxonômicos mais elevados (Kostygov e Yurchenko 2017). Por exemplo, as ambiguidades presentes em alguns táxons definidos pela taxonomia clássica foram mostradas (*Crithidia*, *Blastocrithidia*, *Herpetomonas* e *Leptomonas*), bem como outros reafirmados (*Trypanosoma*, *Leishmania* e *Phytomonas*); foi definida a primeira subfamília, Leishmaniinae, que inicialmente resultou da combinação dos gêneros monoxênicos *Leptomonas* e *Crithidia* com o gênero heteroxênico *Leishmania* (Jirků et al. 2012), além dos táxons que abrigam endossimbiontes, mais tarde definidos nos gêneros *Angomonas* e *Strigomonas* (Maslov et al. 2013); Atualmente, Leishmaniinae juntamente com as subfamílias Strigomonadinae (abriga endossimbiontes) e Blechomonadinae (parasitas de pulgas), bem como alguns gêneros, como *Trypanosoma*, formam a família atual Trypanosomatidae (Yurchenko et al. 2016). Por outro lado, a filogenia molecular demonstrou que o conceito de "tripanossomatídeos inferiores" não era totalmente verdadeiro, uma vez que alguns gêneros como o *Trypanosoma* são basais em relação a alguns gêneros monoxênicos como *Crithidia* e *Blechomonas*, apoiado no fato de que nem todos os

tripanossomatídeos mostram a mesma taxa de evolução. Apesar disso, o parasitismo monoxênico é amplamente aceito como a forma mais primitiva e ancestral de parasitismo, apoiado na posição mais basal do monoxeno *Paratrypanosoma confusum* na filogenia molecular dos tripanossomatídeos (Kaufer et al. 2017). Nesse esquema, acredita-se que os tripanossomatídeos monoxênicos são derivados de bodonídeos de vida livre que se adaptaram ao habitat intestinal de insetos e se tornaram heteroxênicos com o aparecimento de vertebrados na era Mesozóica há 230 M. (Sylvia e Aldo 2019)

Por outro lado, apesar de a filogenia molecular ter melhorado notavelmente a resolução filogenética e ser o guia da taxonomia atual, ainda apresenta algumas imprecisões nas árvores filogenéticas que variam de acordo com o locus ou grupo externo usado, elas também variam dependendo se certos taxa estão incluídos ou não, e também a robustez depende da quantidade de táxons considerados na análise. Uma alternativa para reduzir esses efeitos variáveis é usar múltiplos loci e táxons informativos filogeneticamente. Para este fim, é necessário aumentar o repertório genético de novas espécies e utilizar métodos comparativos com alto poder de resolução que permitam obter árvores filogenéticas o mais exatas possível (Kaufer et al. 2017); isso é fundamental em estudos evolutivos e comparativos.

Com a chegada das novas tecnologias de sequenciamento de alto desempenho, como a genômica e a transcriptômica por tecnologias de segunda geração, passou a haver uma quantidade muito maior de informação genética de organismos tanto de vida livre como parasitas, incluindo monoxenos. Esses dados, por meio de métodos comparativos, como a genômica comparativa e a filogenômica, possibilitaram estudos mais confiáveis e robustos em vários campos como na evolução, epidemiologia, genética e fisiologia dos parasitas. Além disso, permitiram encontrar alvos farmacológicos. Os estudos revelaram vários aspectos genéticos únicos da família Trypanosomatidae com relação ao restante dos eucariotos, bem como homologias e particularidades entre organismos monoxênicos e heteroxênicos.

Os primeiros genomas sequenciados de tripanossomatídeos foram os *Trityp*, que são protozoários hemoflagelados de grande importância médica. Esses genomas, por meio da genômica comparativa, revelaram altos níveis de polaridade de cadeia com grandes conjuntos de genes co-transcritos (transcrição policistrônica) e sintenia conservada entre estas espécies (Berriman et al. 2005; El-Sayed et al. 2005b; Ivens et al. 2005).

No caso de organismos de vida livre, (Opperdoes et al. 2016) analisaram o metabolismo de *Bodo saltans*, protista bacterívoro de vida livre amplamente distribuído em ambientes aquáticos, em conjunto com tripanossomatídeos monoxênicos e dixênicos, a fim de entender a transição do estilo de vida livre para parasitário. Para isso, foram sequenciados e anotados genomas representativos. O número de genes codificantes de proteínas preditas em *B. saltans* foi mais alto que nos tripanossomatídeos, como esperado em um organismo de vida livre. A análise comparativa mostrou

também abundante transferência lateral de genes de origem bacteriana encontrados tanto em “bodonídeos” quanto em tripanossomatídeos; estes genes foram adquiridos pelo ancestral comum de “bodonídeos” e tripanossomatídeos. Estes últimos perderam vias metabólicas completas quando adquiriram estilo de vida parasitário. A presença de um gene codificante de uma enzima de degradação da trealose pode ter fornecido ao ancestral simbiote ou parasita oportunista uma enorme vantagem evolutiva pelo acesso a uma das maiores fontes de energia no inseto. Por outro lado, os tripanossomatídeos adquiriram novos genes que podem ter aumentado a flexibilidade de seus metabolismos e suas possibilidades de adaptação a novos hospedeiros, fazendo dos tripanossomatídeos um dos grupos de parasitas mais bem sucedidos.

O estudo de (Kraeva et al. 2015) sobre o tripanossomatídeo monoxênico *Leptomonas seymouri*, parente muito próximo de *Leishmania* spp., o qual é encontrado frequentemente co-infectando humanos junto com *Leishmania donovani* (Singh et al. 2013), buscou avaliar seu potencial infectivo em vertebrados e sua capacidade de utilizar os vetores de transmissão do parasita dixênico *L. donovani*. Para isto, foram comparados o genoma completo de *L. seymouri* (termoestável a 35 °C) e seus perfis de expressão com aqueles de espécies termosensíveis monoxênicas e termoestáveis dixênicas. Os resultados mostraram que *L. seymouri* tem 26 grupos ortólogos não compartilhados com as outras espécies monoxênicas analisadas, o que poderia ser a causa de sua termoestabilidade. Por outro lado, *L. seymouri* possui dois grupos ortólogos presentes nas espécies dixênicas do gênero *Leishmania* mas não nas outras monoxênicas analisadas: uma proteína kinase-like e uma proteína ubiquinol-citocromo c redutase. Estudos prévios mostraram que as proteínas kinase estão envolvidas na diferenciação de amastigotas em *Leishmania* spp., um processo influenciado por mudanças de temperatura. Se concluiu ainda que *L. seymouri* desenvolveu várias adaptações que permitem que cresça bem a 35 °C, e continua sendo uma espécie predominantemente monoxênica que não consegue infectar macrófagos de mamíferos, seja sozinha ou em coinfeção com *Leishmania*.

Um dos grupos de monoxênicos esquecidos é o gênero *Blechnomonas* (subfamília Blechnomonadinae), que foi descrito pela primeira vez em 2013, com base na análise das sequências dos genes SSU rRNA e gGAPDH, e desde então tem recebido pouca atenção, assim até o momento apenas a sequência genômica de sua espécie-tipo, *Blechnomonas ayalai*, está disponível, porém existem 11 TUs que compõem este gênero, sendo esses : *B. maslovi*, *B. campbelli*, *B. keelingi*, *B. englundii*, *B. juanalfonzi*, *B. lauriereadi*, *B. wendygibsoni*, *B. luni*, *B. danrayi*, *B. pulexsimulantis* (Grybchuk et al. 2018; Votýpka et al. 2013). Esse gênero é composto por organismos extremamente polimórficos, com tamanhos celulares que variam significativamente e apresentam diferentes morfotipos como promastigotas, coanomastigotas e amastigotas, embora outros morfotipos não sejam excluídos. Como consequência, *Blechnomonas* não apresenta características morfológicas únicas que o distingam como gênero (Grybchuk et al. 2018; Votýpka et al. 2013). Esses parasitas são exclusivos das pulgas (Siphonaptera), insetos holometabólicos que se comportam como ectoparasitas de uma

grande variedade de animais, principalmente mamíferos (94%) da ordem Rodentia (74%), além de incluir o homem como hospedeiro habitual, ou comportar-se como vetores para agentes etiológicos de doenças veterinárias e médicas (incluindo vírus, bactérias e protozoários). A maioria das pulgas tem alta especificidade de hospedeiro e algumas coevoluem com elas, embora possa haver trocas acidentais com outros hospedeiros, causando a transmissão de doenças. As pulgas têm ampla distribuição, desde o Ártico até a Antártica, com algumas espécies cosmopolitas, como as relacionadas com humanos, animais domésticos e pecuária, ou com roedores como as espécies *Xenopsylla cheopis*, *Pulex irritans* e *Ctenocephalides felis felis* (Linardi 2017).

As pulgas são colonizadas por *Blechomonas*, principalmente no intestino grosso, posição que está de acordo com o ciclo de vida proposto para esses parasitas. Propõe-se que as larvas das pulgas sejam infectadas pela ingestão de ovos de suas espécies com superfícies contaminadas ou por coprofagia (fezes de pulgas infectadas) e necrofagia (insetos infectados mortos); Desta forma, *Blechomonas* permanece durante todas as fases do desenvolvimento do inseto (metamorfose) até atingir a fase adulta, para então continuar com o ciclo infeccioso. A taxa de infecção por tripanossomatídeos varia amplamente de acordo com as espécies de pulgas, desde espécies sem infecção até aquelas que, além de *Blechomonas*, abrigam uma mistura de outros monoxênicos (*Herpetomonas* e *Leptomonas*) e heteroxênicos (principalmente *Trypanosoma*); portanto, o parasitismo de Siphonaptera por *Blechomonas* não define esse táxon per se, uma vez que outros tripanossomatídeos também exploram esse nicho biológico. Desse modo, devido à mistura infecciosa e ao seu papel como vetor de várias doenças, é possível pensar na possibilidade de o *Blechomonas* ser ocasionalmente transmitido para vertebrados de sangue quente, incluindo o homem (Grybchuk et al. 2018; Votýpka et al. 2013), conforme relatado para outros monoxênicos, embora essa possibilidade ainda seja desconhecida. No entanto, é interessante mencionar a descoberta de uma espécie semelhante a *B. pulexsimulantis* causando uma infecção oportunista em pacientes com HIV (Grybchuk et al. 2018; Pacheco et al. 1998; Votýpka et al. 2013). Isso evidencia mais uma vez a necessidade de estudar esse táxon, assim como os demais táxons monoxênicos, que escondem uma grande quantidade de informações que poderiam redefinir qualquer teoria evolutiva que não leve em conta sua existência (Kaufer et al. 2017; Lukeš et al. 2018; Sylvania e Aldo 2019). Diante dessa necessidade, este estudo tem como objetivo caracterizar o genoma de *Blechomonas* sp. TCC303E, expandindo assim a disponibilidade de dados genéticos neste gênero. Esses dados serão submetidos a uma abordagem de genômica comparativa e filogenômica com outros tripanossomatídeos para um melhor entendimento dos aspectos genéticos e evolutivos do parasitismo em tripanossomatídeos.

Atualmente, estudos filogenéticos posicionam *Blechomonas* como um clado irmão de todos os tripanossomatídeos, com exceção das subfamílias Trypanosomatinae e Paratrypanosomatinae, o que sugere uma origem precoce (Grybchuk et al. 2018). Acredita-se que a adaptação de *Blechomonas*

em pulgas seja um evento raro, provavelmente devido à grande pressão evolutiva exercida pelas peculiaridades do ciclo de vida de seu hospedeiro (Lukeš et al. 2018; McGhee e Cosgrove 1980; Votýpka et al. 2013).

II. Justificativa

Como mencionado acima, o conhecimento dos tripanossomatídeos vem principalmente dos heteroxenos que foram muito mais estudados por terem impactos negativos na sociedade, deixando o maior grupo dos tripanossomatídeos (monoxenos) quase esquecido por muito tempo. Como resultado disso, ainda agora se dispõe de um maior repertório genético dos heteroxenos, como das espécies do gênero *Trypanosoma*, *Leishmania* e *Phytomonas*. Isso gerou vieses e imprecisões nos estudos evolutivos dos tripanossomatídeos que visam entender a origem do parasitismo e a evolução do estilo de vida heteroxeno. Uma alternativa para reduzir essas ambiguidades é aumentar o repertório genético com novas espécies de monoxenos e utilizar métodos comparativos com alto poder de resolução que permitam obter árvores filogenéticas o mais exatas possível.

O uso de novas técnicas tem mudado a taxonomia ao longo do tempo. A chegada e redução dos custos das novas tecnologias de sequenciamento de alto desempenho incrementou de forma significativa a quantidade da informação genética disponível de grupos pouco estudados, como os monoxenos, e tornou possível a aplicação de métodos comparativos de alto poder de resolução que usam milhares de marcadores moleculares, como a genômica e a filogenômica.

Por outro lado, os monoxenos se tornaram relevantes recentemente por serem bons modelos para entender a origem do parasitismo e a evolução dos heteroxenos, devido a serem basais a todos os heteroxenos importantes e compartilharem características genéticas comuns, além de co-infectar os hospedeiros desses heteroxenos que incluem animais domésticos e pacientes com leishmânia ou imunocomprometidos. Essas transmissões incomuns podem nos dar novas perspectivas para entender a origem do ciclo de vida heteroxênico. Em tripanossomatídeos, os nichos biológicos têm um papel importante na sua história evolutiva que tem levado a várias vias de infecção e, em muitos casos, baixa especificidade para seus hospedeiros (parasitas promíscuos); isso resultou na mesma espécie preferindo diferentes locais de colonização e hospedeiros.

Ante à necessidade mencionada anteriormente, este estudo tem como objetivo caracterizar o genoma do monoxeno *Blechnomonas* sp. TCC303E (Classe Kinetoplastea; subfamília Blechnomonadinae) e analisar os aspectos genômicos e evolutivos desse organismo, expandindo assim a disponibilidade de dados genéticos deste gênero, uma vez que, até o momento estava disponível apenas a sequência genômica da espécie-tipo, *Blechnomonas ayalai*. Expandir esses conhecimentos, pode levar a novas e mais robustas hipóteses sobre a origem do parasitismo por comparação com espécies parasitas próximas. Ademais, um dado interessante deste gênero é a descoberta de uma

espécie semelhante a *B. pulexsimulantis* causando uma infecção oportunista em pacientes com HIV.

III. Objetivos

III.1 Objetivo geral

Investigar e comparar o genoma do monoxeno *Blechromonas* sp. TCC303E com os genomas de outros tripanossomatídeos através de técnicas de bioinformática e filogenômica.

III.2 Objetivos específicos

1. Caracterizar os contigs e scaffolds da montagem de *Blechromonas* sp. TCC303E
2. Realizar predição gênica e anotação dos genes de *Blechromonas* sp. TCC303E
3. Comparar o genoma de *Blechromonas* sp. TCC303E com aqueles de outros tripanossomatídeos previamente sequenciados.
4. Verificar, por filogenia multigênica (filogenômica), a posição filogenética de *Blechromonas* sp. TCC303E em relação a outros tripanossomatídeos.

IV. Metodologia e perspectivas

IV.1 Sequências e montagem do genoma

O departamento de Parasitologia do Instituto de Ciências Biomédicas (USP) em conjunto com o Department of Microbiology and Immunology, Virginia Commonwealth University, EUA desenvolveu um grande projeto denominado Assembling the Tree of Life – Phylum Euglenozoa (AToLE). Os dados fornecidos por esse projeto serão essenciais para diversas pesquisas envolvendo o Filo Euglenozoa (Euglenida, Diplonemida e Kinetoplastida), estando o Prof. Dr. João Marcelo Pereira Alves encarregado de estudar os genomas e transcriptomas de diversos organismos da classe Kinetoplastida, além do desenvolvimento de ferramentas de bioinformática. Os dados genômicos novos (não publicados) que foram utilizados neste trabalho vêm, portanto, do projeto AToLE.

O organismo analisado neste projeto, *Blechromonas* sp. TCC303E, foi isolado de uma pulga da espécie *Pulex irritans* (Pulicidae), nos Estados Unidos da América, Estado da Flórida, cidade de Alachua. O genoma do parasita foi sequenciado em nível rascunho através da plataforma Illumina MiSeq, na Virginia Commonwealth University, como parte do projeto mencionado acima.

O sequenciamento do genoma de *Blechromonas* sp. TCC303E gerou um total de 13.967.172 de leituras pareadas (2x300). Com esses dados foi gerada uma montagem utilizando-se o programa

Newbler v. 2.9 (Roche 454), com parâmetros default (exceto pelo uso de `-large`, necessário devido às repetições no genoma, que tornam a montagem lenta).

Também foram baixadas sequências genômicas da base de dados de NCBI (National Center for Biotechnology Information) e da base de dados de Kinetoplastida TriTrypDB (<http://www.tritrypdb.org>, ASLETT et al.,2010). Todos os organismos utilizados são mostrados na seguinte Tabela 1. Além disso, foram utilizadas as sequências dos maxicírculos de espécies de *Leishmania* (*L. tarentolae* (GenBank: M10126.1), *L. braziliensis* (GenBank: LR697134), *L. adleri* (LR697136), *L. guyanensis* (LR697135), *L. infantum* (LR697137), e *L. major* (LR697138) para procurar os *scaffolds* que contenham maxicírculos utilizando a ferramenta de busca por similaridade BLAST v. 2.7.1+ com seu programa BLASTn em modo local e com parâmetros *default*, exceto o E-value com valor de 1e-5.

Tabela 1. Sequências genômicas das 17 espécies de tripanossomatídeos usadas em diversas análises neste estudo.

| Espécie | Tamanho (pb) | contigs/scaffolds | Proteínas | N50 | GC (%) | Base de dados | Versão | |
|---|--------------|-------------------|-------------|--------|------------------------|---------------|-----------|------------|
| <i>Blechnomonas ayalai</i> B08-37 | + | 21.553.048 | 546 (C) | 8.042 | 103.635 bp (L50: 63) | 54,57 | TriTrypDB | 18/01/2017 |
| <i>Crithidia fasciculata</i> CfCI | + | 326.310.40 | 31 (Cr) | 9.489 | 1.318.848 bp (L50: 9) | 57,03 | TriTrypDB | 01/06/2015 |
| <i>Leptomonas pyrrocoris</i> H10 | + | 30.379.976 | 60 (Cr) | 9.878 | 910.096 bp (L50: 9) | 56,44 | TriTrypDB | 30/05/2016 |
| <i>Leptomonas seymouri</i> ATCC30220 | + | 278.017.63 | 1.222 (C) | 8.488 | 70.648 bp (L50: 121) | 55,35 | TriTrypDB | 31/05/2016 |
| <i>Paratrypanosoma confusum</i> CUL13 | + | 27.575.424 | 2.188 (C) | 8.659 | 220.321 bp (L50: 34) | 56,62 | TriTrypDB | 01/06/2016 |
| <i>Endotrypanum monterogei</i> LV88 | ++ | 32.522.248 | 1.961 (Cr) | 8.285 | 885.495 bp (L50: 11) | 51,93 | TriTrypDB | 16/01/2015 |
| <i>Leishmania braziliensis</i> MHOMBR75M2904_2019 | ++ | 33.507.104 | 51 (Cr) | 8.484 | 1.142.536 bp (L50: 10) | 58,04 | TriTrypDB | 30/07/2019 |
| <i>Leishamina donovani</i> BPK282A1 | ++ | 32.444.968 | 36 (Cr) | 8.023 | 1.024.085 bp (L50: 11) | 57,31 | TriTrypDB | 28/05/2016 |
| <i>Leishmania infantum</i> JPCM5 | ++ | 32.803.248 | 36 (Cr) | 8.591 | 1.055.293 bp (L50: 11) | 59,74 | TriTrypDB | 13/07/2018 |
| <i>Leishmania major</i> Friedlin | ++ | 32.855.095 | 36 (Cr) | 8.519 | 1.091.540 bp (L50: 11) | 59,72 | TriTrypDB | 28/05/2016 |
| <i>Leishmania mexicana</i> MHOMGT2001U1103 | ++ | 32.108.741 | 588 (Cr) | 8.246 | 1.044.075 bp (L50: 10) | 59,74 | TriTrypDB | 28/05/2016 |
| <i>Leishmania panamensis</i> MHOMPA94PSC1 | ++ | 30.688.794 | 35 (Cr) | 7.933 | 1.043.456 bp (L50: 10) | 56,25 | TriTrypDB | 05/03/2015 |
| <i>Phytomonas</i> AKH_PRJEB1535 | +- | 17.780.869 | 138 (S) | 6.381 | 428.690 bp (L50: 11) | 43,46 | Genbank | 23/09/2014 |
| <i>Trypanosoma brucei gambiense</i> DAL972 | ++ | 22.148.088 | 11 (Cr) | 7.988 | 2.224.448 bp (L50: 4) | 47,09 | TriTrypDB | 28/05/2016 |
| <i>Trypanosoma cruzi</i> marinkellei B7 | ++ | 38.651.017 | 16.783 (Sc) | 10.228 | 25.781 bp (L50: 297) | 45,12 | TriTrypDB | 29/05/2016 |
| <i>Trypanosoma grayi</i> ANR4 | ++ | 20.934.132 | 2.871 (C) | 10.585 | 16.775 bp (L50: 354) | 53,59 | TriTrypDB | 17/06/2014 |
| <i>Bodo saltans</i> LakeKonstanz | # | 39.864.394 | 2.256 (Sc) | 18.213 | 31.876 bp (L50: 365) | 50,09 | TriTrypDB | 19/11/2015 |

(+) Monoxenos; (++) heteroxenos; (#) organismo de vida livre. (C) *contigs*; (Cr) cromossomas; (S) *scaffolds*; (Sc) super *contigs*.

IV.2 Avaliação da montagem

Existem vários métodos para avaliar uma montagem, mas nenhum é suficiente por si só. Em vista disso, para avaliar a montagem do genoma de *Blechnomonas* sp. TCC303E foram utilizados vários métodos (Gurevich et al. 2013). Primeiro foram procurados contaminantes na montagem utilizando o programa TaxoBlast (<https://sourceforge.net/projects/taxoblast/files/>). Este programa, baseado no programa BLASTx, analisa o contexto genômico para encontrar regiões contaminantes e aquelas originadas por transferência horizontal. O algoritmo divide a montagem em sequências de 1000 bp de comprimento, as quais são comparadas contra uma base de dados local de aminoácidos utilizando o programa de busca por similaridade BLASTx e posteriormente classificar os *scaffolds* em: contaminantes (bactérias), semelhantes a eucariotos e não-classificados; se pelo menos 90% dos hits em uma sequência (de 1000 bp) são de bactérias, a sequência é classificada como contaminante; se for o caso contrário, pelo menos 90% dos hits são contra sequências de eucariotos, a sequência é

classificada como não contaminante (eucariota). Além disso, as sequências que não entram nessas categorias são denominadas como não-classificadas.

Em seguida, foram realizadas comparações com genomas de organismos relacionados disponíveis nas bases de dados públicas, conforme listado na tabela 1. O software utilizado para as comparações foi o pacote de programas MUMmer (Kurtz et al. 2004), especificamente os programas Nucmer para comparações estruturais ao nível de sequências nucleotídicas e PROmer para o nível de aminoácidos; o programa MUMmerplot que gera gráficos do tipo *dotplot* baseado nos alinhamentos gerados por NUCmer e PROmer, o que permite visualizar comparações de montagens e observar alinhamentos entre genomas de espécies relacionadas; e o programa dnadiff que fornece informação detalhada sobre as diferenças entre dois genomas usando o programa NUCmer com parâmetros *default*.

Também foi usado o programa BUSCO (*Benchmarking Universal Single-Copy Orthologs*) v4 (Simão et al. 2015) para avaliar a integridade do genoma, mediante a comparação do número de genes centrais preditos nesta montagem contra os de organismos do filo Euglenozoa armazenados na base de dados euglenozoa_db10 (Versão de data 2019/11/27), o qual fornece medidas quantitativas relacionadas a essa comparação.

Além disso, medidas estatísticas descritivas da montagem mencionada acima foram examinadas em mais detalhes para avaliar a contiguidade das sequências montadas. Programas como mfsizes (<https://sourceforge.net/projects/mfsizes>) calculam número de sequências, tamanhos médios e N50/L50 (assim como medidas relativas a 90 e 95% do conteúdo total de sequência), conteúdo de GC para cada sequência ou média ponderada geral, entre outras medidas. A avaliação destes números ajudou tanto na avaliação da qualidade da montagem quanto a identificar eventuais problemas na montagem e presença de sequências contaminantes (por exemplo, aquelas que são longas mas apresentam conteúdo GC muito diferente da média).

IV.3 Predição gênica

Baseado nos estudos prévios dos Kinetoplastida, foi assumida uma organização do genoma de *Blechnomonas* sp. TCC303E semelhante àquela vista nos genomas de outros Kinetoplastida (ou seja, quase sem nenhum intron), por isso a predição gênica foi feita utilizando ferramentas desenvolvidas para genomas bacterianos. Assim sendo, utilizamos ferramentas como Glimmer3 (Delcher et al. 2007), que utiliza modelos de Markov interpolados (IMMS) para identificar as regiões codificadoras e distingui-las do DNA não-codificante. Também foi utilizado GeneMarkS v. 4.28 para comparar resultados com Glimmer3. Esta versão do GeneMarkS é uma extensão do original (Besemer et al. 2001) e permite prever genes em organismos eucarióticos sem introns, utilizando modelos ocultos de Markov (HMM) iterativos num processo de treinamento não supervisionado.

Genes para RNAs ribossômicos e transportadores foram preditos usando os programas RNAmmer (Lagesen et al. 2007) e tRNAscan-SE (Lowe e Eddy 1997), respectivamente. Outros RNAs não codificantes (ncRNA) foram preditos com o uso do programa Infernal (Nawrocki e Eddy 2013), usando o banco de dados Rfam (Griffiths-Jones et al. 2003).

IV.4 Anotação

Para a anotação dos modelos gênicos preditos (CDS) na etapa anterior, primeiramente foram comparados contra a base de dados NR (Non-redundant) do NCBI (National Center for Biotechnology Information), utilizando a ferramenta de busca por similaridade BLAST v. 2.7.1+ com seu programa BLASTp em modo local e com parâmetros *default*, exceto o *E-value* com valor de $1e-5$. Os melhores *hits* foram filtrados no ambiente linux com comandos bash.

Os modelos gênicos preditos também foram anotados no nível protéico, por busca de domínios na base de dados PFAM v.31 (Punta et al. 2012), utilizando a ferramenta HMMER (3.1b2), especificamente com o programa hmmscan com parâmetros *default*, exceto o *E-value* com valor de $1e-6$. HMMER utiliza HMM para procurar similaridades baixas derivadas de homologies remotas com a maior sensibilidade possível (Johnson et al. 2010).

Para a anotação metabólica dessas sequências utilizamos o ASGARD (Alves e Buck 2007), que tem como função principal construir vias metabólicas, mas também utilizamos a anotação geral para fazer comparação usando termos de Gene Ontology, KEGG Orthology e Enzyme Commission. Também foram utilizadas outras plataformas automatizadas de anotação como o eggNOG 4.5 (Huerta-Cepas et al. 2016), base de dados pública de grupos ortólogos (GOs) de proteína e anotações funcionais em diferentes níveis taxonômicos e que anota funcionalmente com base nos grupos ortólogos, usando o algoritmo de agrupamento não supervisionado baseado em grafos. O eggNOG gerou anotações classificadas nas categorias funcionais de KOGs (Eukaryotic Orthologous Groups), nos termos de Gene Ontology (GO) e KEGG (KO).

IV.5 Genômica comparativa

Com a genômica comparativa podemos analisar e comparar genomas de diferentes espécies com o propósito de obter um melhor entendimento de como estes evoluíram, assim como determinar a função de genes e de regiões não codificantes do genoma. Dessa maneira, além das comparações cromossômicas mencionadas acima em "Avaliação da montagem", foi utilizado o programa OrthoFinder (Emms e Kelly 2019) para a identificação de grupos de genes ortólogos (necessários para as análises filogenéticas e de conteúdo gênico).

O OrthoFinder utiliza o método dos melhores hits bidirecionais (*best bidirectional hits*),

associados a uma etapa final de markov *clustering*, sendo muito mais rápido que os métodos baseados em análises filogenéticas completas e mais facilmente automatizado. Relações de ortologia inferidas pelo OrthoFinder foram confirmadas durante as análises filogenéticas posteriores.

Os dados gerados pelo ASgard, descritos acima, também servem para análise metabólica comparativa, permitindo a comparação simultânea, em cada via metabólica, de até cinco organismos diferentes.

A perspectiva desta etapa foi a obtenção de um levantamento de quanto *Blechomonas* sp. TCC303E difere de outros cinetoplastídeos, em especial seus parentes mais próximos, e quanto têm em comum, tanto em termos de conteúdo gênico quanto intergênico.

IV.6 Filogenômica

A filogenômica é uma sofisticada abordagem evolutiva que aproveita as áreas de genômica e filogenia para estudar genomas completos ou grande parte deles e é também uma abordagem promissora para resolver a árvore da vida. A filogenômica permite obter uma melhor resolução do que pode a filogenia através da geração de uma árvore de espécie mais robusta. Assim, neste estudo, os dados obtidos nas etapas de anotação e comparação do genoma de *Blechomonas* sp. TCC303E com os outros tripanossomatídeos foram analisados através do ponto de vista da filogenômica para compreender a relação evolutiva do organismo dentro dos tripanossomatídeos.

Para obter a árvore de espécie se seguiu uma abordagem que consistiu primeiro na geração dos alinhamentos dos grupos ortólogos de cópia única através do programa MAFFT (Kato e Standley 2013) e as regiões de baixa qualidade dos alinhamentos foram filtradas com o programa TRIMAL (-gt 0.9 -cons 60) (Capella-Gutiérrez et al. 2009). Depois, esses alinhamentos foram usados para gerar árvores de genes utilizando a metodologia de máxima verossimilhança com o programa IQ-TREE (Nguyen et al. 2015). Por último, a árvore das espécies foi inferida a partir das árvores dos genes sob o modelo de coalescência de múltiplas espécies com o programa ASTRAL (Zhang et al. 2017). Finalmente, a filogenia foi visualizada com a ferramenta FigTree e editada com Inkscape.

V. RESULTADOS E DISCUSSÃO

V.1 Características e avaliação da montagem genômica

O genoma de *Blechomonas* sp. TCC303E foi sequenciado em nível rascunho, através da

plataforma Illumina MiSeq, resultando num total de 13.967.172 leituras pareadas (2 x 300), totalizando 4.190.151.600 bases. A montagem resultou em um tamanho total de ~22,2 Mbp (Tabela 2), com uma cobertura média de 38,6, e sua sequência está distribuída em 728 *scaffolds*, com um valor N50 de 61.737 bp (~61,7 Kbp) (L50 = 110) -- o que significa que pelo menos 50% da sequência se encontra nos 110 maiores *scaffolds* (o menor dos quais tem ~61,7 kbp de comprimento). O tamanho do genoma montado é comparável ao resto dos genomas usados neste estudo, e é próximo ao de *B. ayalai* (21,5 Mbp).

Tabela 2. Principais estatísticas da montagem do genoma de *Blechromonas* sp. TCC303E.

| Característica | <i>Blechromonas</i> sp. TCC303E |
|----------------------------------|---------------------------------|
| Medida da montagem (Mb) | 22.286.162 |
| Número de contigs | 728 |
| Total Ns | 459.580 |
| N50 | 61.737 bp (L50: 110) |
| Média GC (%) | 50,3 |
| Porção codificante do genoma (%) | 59,6 |
| Número total de pb | 22.286.162 |

O menor comprimento dos *scaffolds* montados tem 2 Kbp e o maior tem 214.616 bp (~2,1 Mbp). A composição GC média do genoma é de 50,30%, o que é um dos menores em comparação com os outros tripanossomatídeos (Tabela 1) neste estudo, sendo maior apenas que as médias das espécies *T. brucei*, *Phytomonas*, e *T. cruzi*, e também incluindo *B. ayalai*, a única espécie do gênero *Blechromonas* disponibilizada na base de dados TriTrypDB, com uma média de 54,57% de GC. A média de GC é mais próxima daquela da espécie *E. monterogeii* (51,9%) e quase igual ao do organismo de vida livre *B. saltans* (50,09% de GC). Esses resultados parecidos em tamanho e número de CDS predito com a espécie mais próxima indicam uma boa qualidade da montagem e da predição do número de CDS, mas também foram feitas outras análises para avaliar a qualidade desses processos. Além disso, esses dados vão de acordo com a gama de variação dos comprimentos dos genomas e conteúdo de CDS observados em outros tripanossomatídeos (tabela 1), o que agrega a ideia de que os organismos parasitas têm genomas comparáveis, em comprimento e conteúdo gênico, com eucariotos modelo de vida livre (Jackson et al. 2016).

Uma das características compartilhadas pelos tripanossomatídeos é o cinetoplasto (kDNA), com maxicírculos com comprimentos que oscilam entre 20 e 40 kbp, contendo porcentagens de GC que variam dependendo da espécie, oscilando entre 18 e 42%, e contendo genes envolvidos na geração de energia (incluindo subunidades de citocromo oxidase, NADH desidrogenase e a ATP sintase) e na codificação de rRNAs e tRNA (Jensen e Englund 2012; Kaufer et al. 2019). Dentre os *scaffolds* da montagem, foi encontrado apenas um *scaffold* com um comprimento de ~18,2 Kbp (com 2.795 Ns) que contém DNA mitocondrial e com uma composição de GC de 19,7%, distante da média

da montagem (50,30%). Ainda, foram procurados os maxicírculos no genoma de *B. ayalai* para comparação, sendo também encontrado DNA mitocondrial em um único pedaço de sequência genômica (*contig*) com uma composição de 22,08% de GC e um comprimento de ~16,9 Kbp (com 83 Ns). Assim, foram encontrados em nossa montagem dez genes que codificam para subunidades de NADH desidrogenase (Tabela 3).

Tabela 3. Genes encontrados no maxicírculo de *Blechnomonas* sp. TCC303E usando a ferramenta BLASTx contra o banco de dados NR, e espécies com genes mais semelhantes.

| Gene | Espécie |
|--|------------------------------|
| NADH dehydrogenase subunit 5 | <i>Strigomonas oncopelti</i> |
| NADH dehydrogenase subunit 5 | <i>Strigomonas oncopelti</i> |
| NADH dehydrogenase subunit 5 | <i>Leishmania donovani</i> |
| NADH dehydrogenase subunit 5 | <i>Leishmania donovani</i> |
| NADH dehydrogenase subunit 5 | <i>Angomonas deanei</i> |
| NADH dehydrogenase subunit 5 | <i>Angomonas deanei</i> |
| NADH2 dehydrogenase (ubiquinone) chain 5 | <i>Leishmania tarentolae</i> |
| NADH2 dehydrogenase (ubiquinone) chain 5 | <i>Leishmania tarentolae</i> |
| NADH dehydrogenase subunit 5, partial | <i>Leptomonas pyrrocoris</i> |
| NADH dehydrogenase subunit 5, partial | <i>Leptomonas pyrrocoris</i> |

No mais, foram procuradas, com o programa Taxoblast, sequências contaminantes que pudessem ter sido incluídas nos *scaffolds* durante o processo de montagem do genoma de *Blechnomonas* sp. TCC303E. Os resultados mostram que só um *scaffold* (Ble_1_s00696) foi classificado como contaminante (bactéria); tem 2.437 bp de comprimento, o que representa 0,011% da montagem. Desse *scaffold*, a parte contaminante representou 39,76 %, portanto, com o fim de não perder material genético do próprio organismo em estudo, por exemplo sequências curtas não transcritas ou não codificantes como moléculas de RNA não codificantes, decidiu-se não o remover da montagem. Os *hits* recuperados para o *scaffold* foram proteínas hipotéticas das bactérias *Grimontia celer*, *Lactobacillus sanfranciscensis* e *Pseudomonas aeruginosa*. Os *scaffolds* classificados como não contaminantes foram 573 e os não classificados foram 154. Os *scaffolds* destas duas categorias foram filtrados para deixar só aqueles que tenham pelo menos um *hit* de um organismo kinetoplastida utilizando um *script* feito *in-house*, obtendo-se só um *scaffold* (Ble_1_s00553) da categoria dos não classificados que não tem *hit* de organismos Kinetoplastida, com um comprimento de 5.385 bp (0,024% da montagem) e com 6 *hits* que pertencem a vírus, bactérias e eukaryota, mas cobrem apenas 18,57% do comprimento total do *scaffold*, optando-se também por não removê-lo.

Por outro lado, a montagem foi avaliada em termos de conteúdo esperado de genes (a completude da montagem) pelo programa BUSCO usando o banco de dados para Euglenozoa. Os conjuntos de genes esperados são denominados categorias BUSCO e são determinados conforme a informação evolutiva dos genes ortólogos de cópia única, quase universais. Os resultados são resumidos na figura 1.

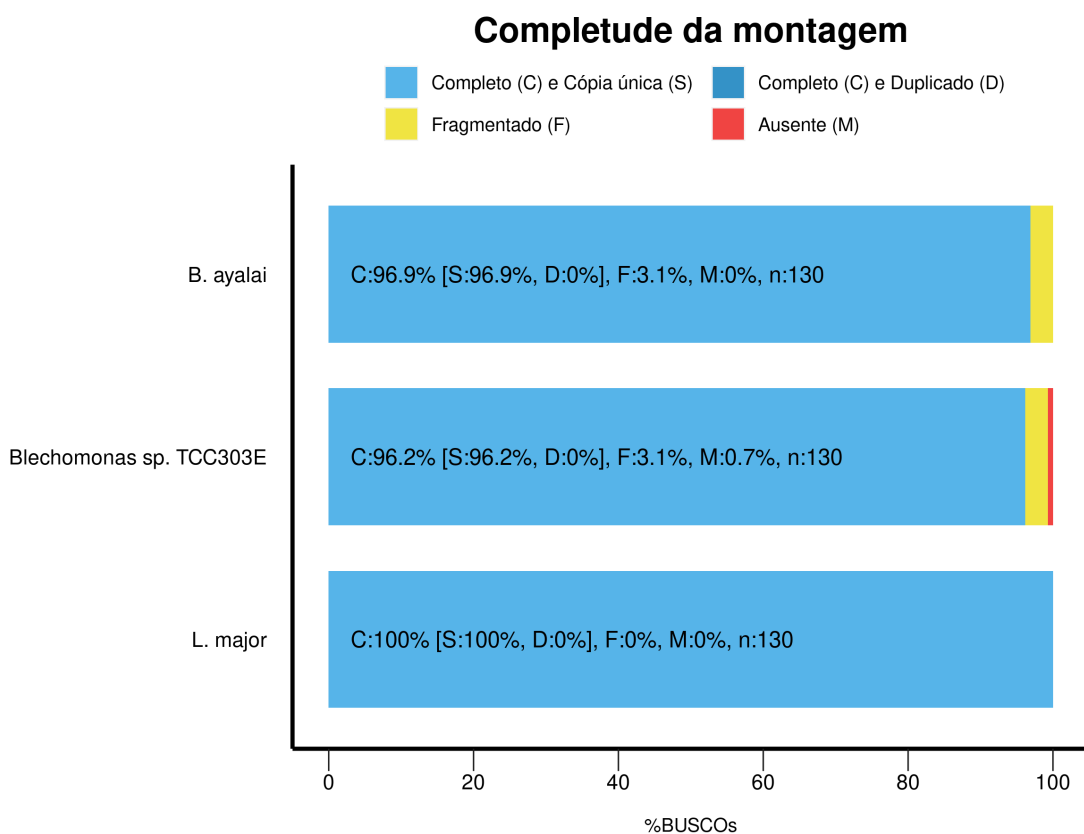


Figura 1. Comparação das categorias BUSCO para avaliar a completude da montagem de *Blechomonas sp. TCC303E* com respeito aos genomas de *B. ayalai* e *Leishmania major*. No eixo horizontal estão as porcentagens BUSCO e no eixo vertical estão as espécies, além das legendas das cores na parte superior.

Para o genoma em estudo, os resultados mostram apenas 8 (0,7%) categorias BUSCO ausentes e um valor alto de completude, com 125 (96,2%) das 130 categorias BUSCO, além de apresentar valores quase iguais comparados com *B. ayalai* e muito próximos aos de *L. major*, que tem todos os genes esperados (Figura 1). Assim, os resultados sugerem que a montagem é de alta qualidade, comparado com testes em organismos modelos com montagens de alta qualidade que têm valores maiores que 95% de completude (Seppey et al. 2019).

Além das medidas mencionadas acima, a montagem do genoma de *Blechomonas sp. TCC303E* foi comparada com o genoma de *B. ayalai* usando os programas NUCmer e PROmer para comparações de sequências nucleotídicas e de aminoácidos, respectivamente, sendo visualizados com o programa MUMmerplot, conforme explicado na seção dos métodos. Os resultados das semelhanças e diferenças estruturais são apresentados nas figuras 2 e 3.

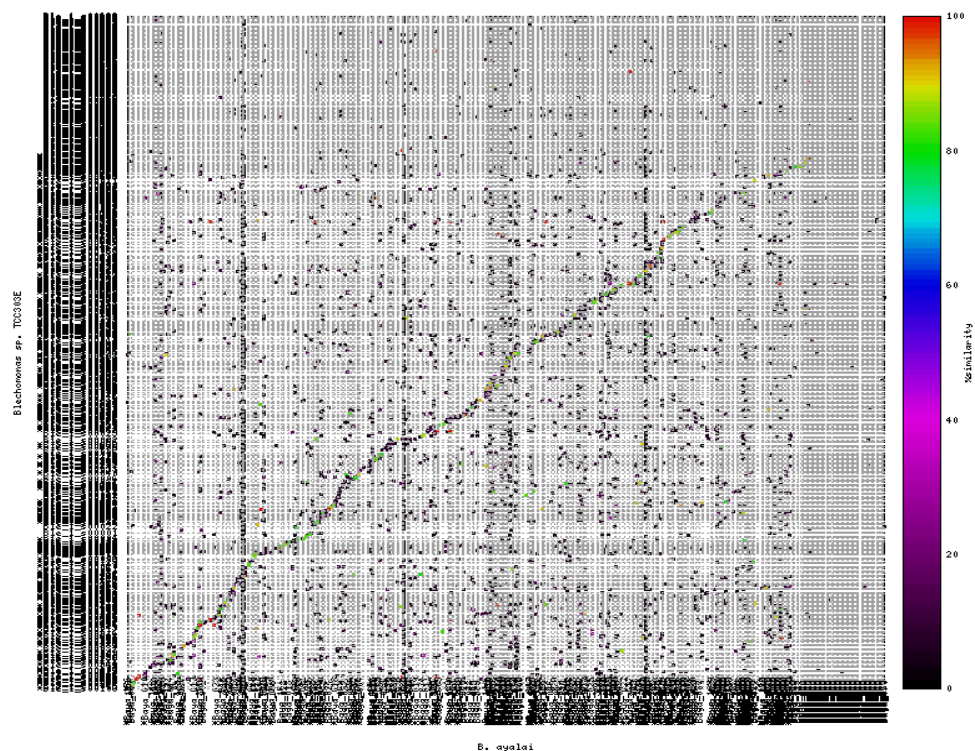


Figura 2. Gráfico *dotplot* gerado pelo pacote MUMmerplot do resultado do programa NUCmer para a comparação estrutural usando sequências nucleotídicas da montagem de *Blechomonas* sp. TCC303E com respeito ao genoma de *B. ayalai*. No eixo horizontal estão distribuídos todos os *scaffolds* (546) do genoma de referência, *B. ayalai*, e no eixo vertical estão todos os *scaffolds* (728) da montagem do genoma de *Blechomonas* sp. TCC303E. As cores representam porcentagens de similaridade nos alinhamentos. A diagonal mostra a alta sintenia entre os genomas comparados.

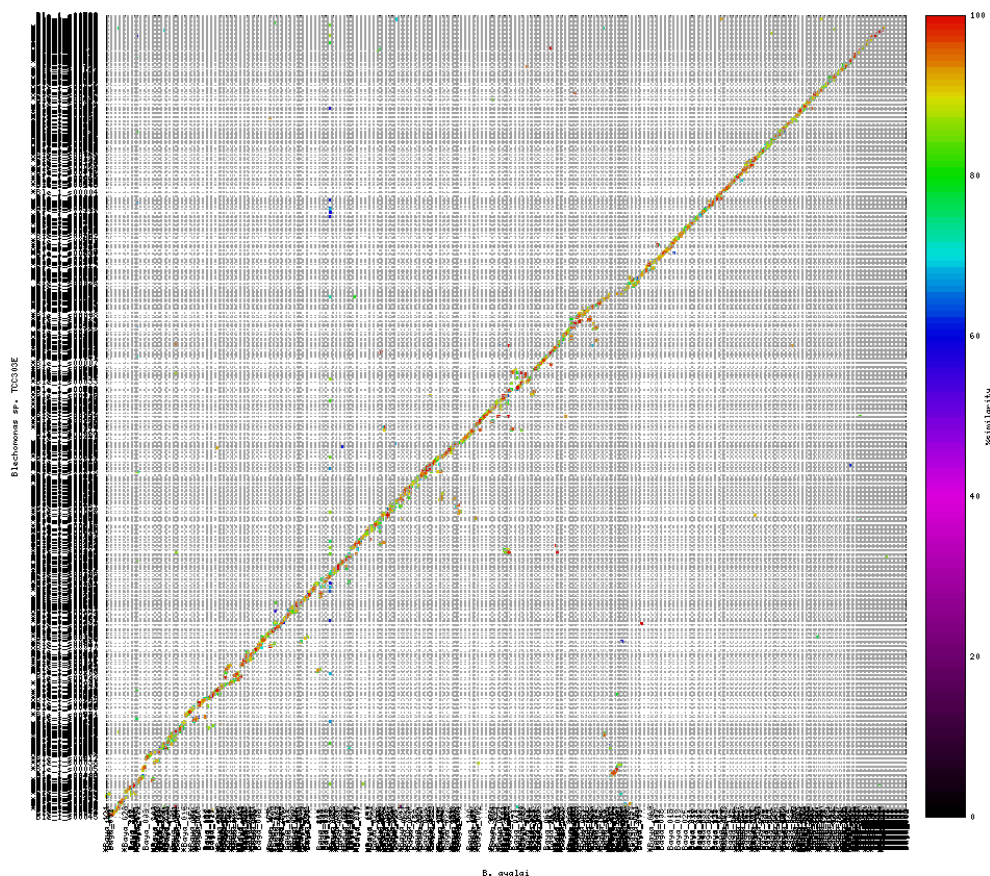


Figura 3. Gráfico *dotplot* gerado pelo pacote MUMmerplot do resultado do programa PROmer visando a comparação estrutural no nível de seqüências de aminoácidos da montagem de *Blechomonas* sp. TCC303E com o genoma de *B. ayalai*. No eixo horizontal estão distribuídos todos os *scaffolds* (546) do genoma de referência, *B. ayalai*, e no eixo vertical estão todos os *scaffolds* (728) da montagem do genoma de *Blechomonas* sp. TCC303E. As cores representam porcentagens de similaridade nos alinhamentos. A diagonal mostra a alta sintenia entre os genomas comparados.

A figura 2 mostra que a comparação de nucleotídeos tem alta cobertura em similaridade entre as seqüências dos organismos comparados, onde foram alinhados 465 *scaffolds* de *Blechomonas* sp. TCC303E. Os alinhamentos são em grande parte de baixa similaridade, o que sugere uma alta divergência no nível nucleotídico conforme representado na diagonal principal do gráfico *dotplot*.

Paralelamente, foi feita uma comparação de aminoácidos (figura 3), que é um enfoque mais sensível do que o anterior para procurar regiões conservadas mais divergentes, visto que no nível protéico a seqüência é mais conservada (Koonin e Galperin 2003). O alinhamento no nível das proteínas candidatas tornou possível a melhor visualização dos valores de similaridade das seqüências relacionadas, sendo possível ainda, uma melhor cobertura da montagem de *Blechomonas* sp. TCC303E em relação a *B. ayalai*, como se observa no gráfico. Outra análise do mesmo tipo foi feita

usando como referência o genoma de *L. major*, o que forneceu resultados concordantes (Anexo A).

V.2 Predição de genes codificadores de proteínas e RNAs.

Após a montagem e avaliação do genoma, a seguinte etapa é a predição das possíveis regiões codificadoras dos diversos elementos do genoma, como as sequências que codificam proteínas (CDS, *Coding DNA sequence*) e sequências que transcrevem RNAs não-codificantes (tRNAs, rRNAs, e ncRNAs). Esta etapa é fundamental, sendo o início da predição das características genéticas dos genomas (estrutura e repertório genético). Nesta etapa foram utilizados diversos programas de predição *ab initio* baseados em modelos ocultos de Markov (HMM) e modelos de covariância (CV) para descrever as estruturas lineares e estruturas secundárias dos genes, respectivamente. No caso da predição de genes codificadores de proteínas foram utilizados os programas Glimmer e GeneMarkS (que usam HMM), os quais geraram 26.672 e 8.044 modelos gênicos (CDS), respectivamente. Os programas foram utilizados com parâmetros “*default*” e com código genético padrão; ademais, no caso de GeneMarkS foi usada a opção “*Intronless eukaryotic*” para organismos eucariotos com pouco ou nada de introns.

As CDS obtidas pelo programa GeneMarkS foram selecionadas para a etapa posterior de anotação, já que são semelhantes aos dados do genoma de *B. ayalai* no número de CDS (8.042) e conteúdo médio de GC dos CDS (54,57%) (Tabela 1). Se comparados com a predição de Glimmer, os resultados de GeneMarkS foram mais realistas, provavelmente por causa do seu algoritmo iterativo que incorpora sítios de ligação aos fatores de transcrição, o que melhora a acurácia para identificar os inícios dos genes, como foi testado por (Besemer et al. 2001), onde comparou esse programa contra Glimmer em organismos procariotos. Glimmer predisse mais CDS, mas seria irrealista considerar que uma espécie próxima e com mesmo tamanho aproximado de genoma possa ter mais de quatro vezes o número de CDS, ademais se observou que os CDS têm comprimentos mais reduzidos e podem ser falsos positivos dado que o programa usa o código genético para bactérias.

Da mesma forma que para a identificação dos genes que codificam para proteínas, diferentes classes de RNAs não codificantes (ncRNA) foram identificadas através de ferramentas específicas que além das estruturas lineares dos genes (usando HMM) consideram as estruturas secundárias na forma de modelos de covariância (CV). Com os resultados gerados pelo programa RNAmmer, observaram-se somente duas classes de RNAs ribossômicos preditos (subunidades 28S e dois 8S) o que levou à busca da subunidade 18S nos *contigs* usando o mesmo programa, onde tampouco foram encontradas sequências 18S. Isto levou à procura dessa sequência mediante o programa BLASTn, utilizando a sequência 18S de *B. ayalai* (de 2.199 pb), desta forma foi encontrado um *hit* com 98,2% de identidade e 961 pb de comprimento de alinhamento. Pelo contrário, o baixo desempenho do

programa RNAmmer para identificar sequências 18S em procariotos e eucariotos já foi observado em seu artigo, devido a uma grande divergência global que apresentam estas sequências que dão lugar a estados de inserção nos HMM e desta forma dificultam a visualização destes genes (Lagesen et al. 2007). Por outro lado, o comprimento mais curto do gene 18S nos *contigs* e sua completa ausência nos *scaffolds* é possivelmente causado por erros durante o processo da montagem em *contigs* e posterior eliminação quando formados os *scaffolds*.

Nos resultados do programa tRNAscan-SE observaram-se 71 genes para tRNAs de transferência (mais um pseudogene), abrangendo a ligação de todos os aminoácidos, sendo o maior número de cópias para alanina (7) e arginina (7), ácido glutâmico (6) e leucina (6) –ver detalhes no Anexo B. Outros RNAs não codificantes responsáveis pela maturação do mRNA foram preditos pelo pacote Infernal 1.1. Um total de 62 genes para outros RNAs não codificantes foram encontrados, sendo estes próximos em números ao do genoma de *B. ayalai* que tem 76 genes, mas bem diferente dos outros genomas (Tabela 4).

Uma das categorias mais comuns de ncRNAs encontradas em *Blechnomonas* sp. TCC303E foi a dos pequenos RNAs nucleolares (em inglês, snoRNAs), os quais fazem importantes modificações nos rRNAs dos tripanossomatídeos para se adaptarem melhor durante os ciclos de infecção (Rajan et al. 2019). Os snoRNAs representam 39,7% do total predito (24/62), e dentro deste grupo estão os snoRNAs que pertencem especificamente a tripanossomatídeos e são 20, ou 83,3% (20/24) (próximo do encontrado em *B. ayalai* (28), onde encontramos os seguintes snoRNAs: TB10Cs3H1, TB10Cs3H2, TB9Cs2H1, TB10Cs2H1, TB10Cs1H1, TB10Cs1H3, TB10Cs1H2, TB11Cs4H1, TB11Cs4H2, TB10Cs5H3, TB8Cs3H1, TB8Cs3H1, TB9Cs1H2, TB9Cs1H1, TB9Cs1H1, TB9Cs1H3, TB10Cs4H2, TB10Cs4H4, TB9Cs3H1, TB9Cs3H2. Outra categoria mais abundante, e que tem uma função na interação parasita-hospedeiro, é a categoria dos microRNAs (Linhares-Lacerda e Morrot 2016; Zheng et al. 2013), os quais representam 24,2% (15/62) dos ncRNAs.

Tabela 4: Resumo das características do genomas de 10 espécies de tripanossomatídeos.

| Característica | B.p ** | B.a | C.f | L.p * | P.c | E. m | L. ma * | P. AKH | T. b * | B. s |
|------------------------------------|---------|---------|--------|---------|-----------|-----------|---------|-----------|--------|------------|
| Número de sequências | 728 (S) | 546 (C) | 31(Cr) | 60 (Cr) | 2.188 (C) | 1.961(Cr) | 36 (Cr) | 138(S) | 11(Cr) | 2.256 (Sc) |
| Medida da montagem (Mb) | 22.3 | 21.5 | 32.6 | 30.4 | 27.6 | 32.5 | 32.9 | 17.8 | 22.1 | 39.9 |
| Conteúdo GC de CDS (%) | 50,3 | 54,57 | 57,03 | 56,44 | 56,62 | 51,93 | 59,72 | 43,46 | 47,09 | 50,09 |
| Conteúdo N | 459.580 | 143.744 | 851 | 113.719 | 2.286.691 | 515.994 | 13 | 1.679.117 | 37.367 | 1.308.364 |
| Genes codificadores de proteínas | 8.044 | 8.042 | 9.489 | 9.878 | 8.659 | 8.285 | 8.519 | 6.381 | 7.988 | 18.213 |
| Genes para ncRNAs (*) | 62 | 76 | 171 | 189 | 189 | 316 | 409 | 39 | 136 | 100 |
| Genes para RNAs de transferência | 71 | 69 | 104 | 103 | 59 | 104 | 82 | 65 | 62 | 56 |
| Genes para RNAs ribossômico | 3 | 3 | 3 | 5 | 1 | 6 | 23 | 1 | 13 | 3 |
| Pequenos RNAs nucleolares (snoRNA) | 24 | 31 | 146 | 167 | 15 | 81 | 245 | 18 | 102 | 19 |

Blechnomonas sp. TCC303E (B.p), *B. ayalai* (B.a), *C. fasciculata* CfCl (C.f), *Le. pyrrocoris* H10 (L.p), *P. confusum* CUL13 (P.c), *E. monterogeei* LV88 (E.m), *L. major* Friedlin (L.ma), *Phytomonas* AKH_PRJEB1535 (P.AKH), *T. brucei* gambiense DAL972 (T.b), *B. saltans* Lake Konstanz (B.s). (*)

Apenas uma espécie de cada gênero foi escolhida. (**) CDS preditos com GeneMarkS, rRNA com RNAmmer, tRNA com tRNAscan-SE, ncRNAs com infernal. *Contigs* (C), *scaffolds* (S), cromossomos (Cr), Super *contigs* (Sc).

V.3 Anotação funcional

Com as potenciais sequências de proteínas preditas, o próximo passo foi obter um catálogo de possíveis funções biológicas derivadas de buscas de homólogos putativos por similaridade de sequência e dos ortólogos com o *Reciprocal best bidirectional hits* (rbbh). Uma fonte rica de anotações funcionais é derivada da comparação com sequências já anotadas de outros organismos armazenadas nos bancos de dados. Para essas comparações, métodos baseados em busca por similaridade de sequência são geralmente a primeira escolha de abordagem, os quais são particularmente eficazes em genomas sem íntrons e com alta densidade gênica, como é hipotetizado para o nosso genoma em estudo. O método mais usado nestas buscas por similaridade é o algoritmo BLAST devido a sua velocidade, seletividade alta e flexibilidade (Koonin e Galperin 2003). Em vista disso, neste trabalho foi usado o BLAST, especificamente, foi usada a ferramenta BLASTp contra o banco de dados NR do NCBI (Pruitt et al. 2005), baixados em agosto de 2020.

Esta abordagem resultou na anotação de 6.218 sequências putativas de proteínas das 8.044 sequências proteicas preditas com o programa GeneMarkS (com um mínimo de 50% de cobertura da *query* e 30% de porcentagem de identidade), representando 77,3%, das quais 44,6% (2.772/6.218) são anotadas como proteínas hipotéticas. O maior número de *best hits* veio do gênero *Leishmania*, envolvendo 10 espécies, que representam 26,7% (1.662), e outros gêneros com maiores porcentagens de anotações foram: *Leptomonas* com 23,3% (1.446), *Trypanosoma* com 22,8% (1.417), e *Phytomonas* com 14,2% (880). Da mesma forma, no nível de espécie (depois dos isolados do gênero *Phytomonas* sp. que representam 14,2% do total anotado (6.218), as maiores porcentagens de *best hits* foram obtidas das seguintes espécies: *Leptomonas pyrrocoris* com 12,1% (755), *Leptomonas seymouri* com 11,1% (691), e *Strigomonas culicis* com 6,5% (401). Essas porcentagens de similaridade coincidem com o posicionamento filogenético do gênero *Blechnomonas* nos tripanossomatídeos, onde *Blechnomonas* é um clado irmão da subfamília Leishmaniinae e próximo ao clado das espécies *Trypanosoma* spp. Observam-se nesse contexto expansões e contrações dos repertórios genéticos desses organismos (Lukeš et al. 2018), eventos evolutivos que podem ter influenciado nas porcentagens de hits obtidas.

Paralelamente à busca por similaridade dos candidatos a homólogos, foi feita uma atribuição de funções por busca de ortólogos por meio da técnica rbbh, com limiares de 50 e 30% para cobertura da *query* e porcentagem de identidade, respectivamente, usando as proteínas preditas de *Blechnomonas* sp. TCC303E e as proteínas preditas de *B. ayalai*. Com os resultados de rbbh obteve-se um incremento das proteínas anotadas, com 7.002 *hits*, que representam 87% do total das proteínas

preditas pelo GeneMarkS, as quais apresentam também um incremento no número de proteínas hipotéticas 57% (3.991). Essa abundância no número de proteínas hipotéticas é um problema recorrente em projetos genômicos, como são particularmente observados em genomas de tripanossomatídeos, onde o número de genes anotados como codificantes de proteínas hipotéticas geralmente é maior que a metade, como por exemplo em *L. major* (68% de proteínas hipotéticas) (Callejas-Hernández et al. 2018; Grisard et al. 2010; Silber e Pereira 2012; Tschoeke et al. 2014). Por exemplo, a base de dados TriTrypDB apresenta 552.521 genes, dos quais 238.956 (43,2%) genes codificam proteínas hipotéticas (consultado em setembro de 2020).

V.4 Análise de sintenia de *Blechnomonas* sp. TCC303E

Conforme observado na etapa anterior, *Blechnomonas* sp. TCC303E compartilha um maior grau de similaridade com outros tripanossomatídeos conforme as distâncias filogenéticas são mais curtas. Assim, a espécie *B. ayalai* atingiu a maior porcentagem de similaridade (87%), além de um número semelhante de genes preditos. Além dessas semelhanças genéticas, esperava-se que compartilhassem um maior grau de conservação de sintenia e também altos graus de sintenia com os outros tripanossomatídeos, uma vez que é conhecido que os tripanossomatídeos possuem uma arquitetura genômica semelhante, conformada por blocos sintênicos altamente conservados, inclusive entre grupos de tripanossomatídeos distantes como nos genomas TriTryp, com uma sintenia de 94% entre os genes conservados, e que entre espécies do mesmo gênero podem superar os 99% (Flegontov et al. 2016; Schmid-Hempel et al. 2018; Sloan et al. 2019; Tschoeke et al. 2014). Para realizar a análise de sintenia foram escolhidas as espécies melhor estudadas, sendo escolhida a espécie mais próxima (*B. ayalai*) e uma mais distante (*L. major*). Para tal propósito foi usado o pacote de software SyMAP v5.0.6 que faz uso do pacote MUMmer para detectar blocos sintênicos imperfeitos (referidos como blocos sintênicos), que são regiões gênicas conservadas independentemente da ordem ou fita em que são localizadas. O SyMAP foi executado com parâmetros *default*, exceto para considerar sequências com tamanho mínimo de 100 Kbp, um valor próximo e que abrange o grupo que compõe o L50 dos genomas comparados.

O SyMAP filtrou as sequências com comprimentos superiores ao limiar e aqueles com ocorrência de sintenia, resultando em 46 *scaffolds* de *Blechnomonas* sp. TCC303E, 66 *contigs* de *B. ayalai* e 2 cromossomos de *L. major*. Neste último organismo, suas sequências não foram filtradas pelo limiar, pois os comprimentos dos cromossomos são maiores que o limiar. A partir da análise com SyMAP, foram obtidos 39 blocos sintênicos na comparação com *B. ayalai* (que abrigam 374 genes anotados de *B. ayalai*), dos quais 27 são menores do que 100 Kbp e 12 blocos sintênicos, os quais, têm comprimentos entre 100 Kbp e 1 Mbp. Ademais, dos 39 blocos sintênicos, 22 são inversões (Anexo C). Em comparação com *L. major* foram obtidos 7 blocos sintênicos, com 5 blocos menores do que 100 Kbp e 2 estão com comprimentos entre 100 Kbp e 1 Mbp. Este número menor de blocos

sintênicos em *Blechnomonas* sp. TCC303E quando comparado com *L. major* está correlacionado com a distância filogenética entre esses organismos e os tamanhos dos scaffolds.

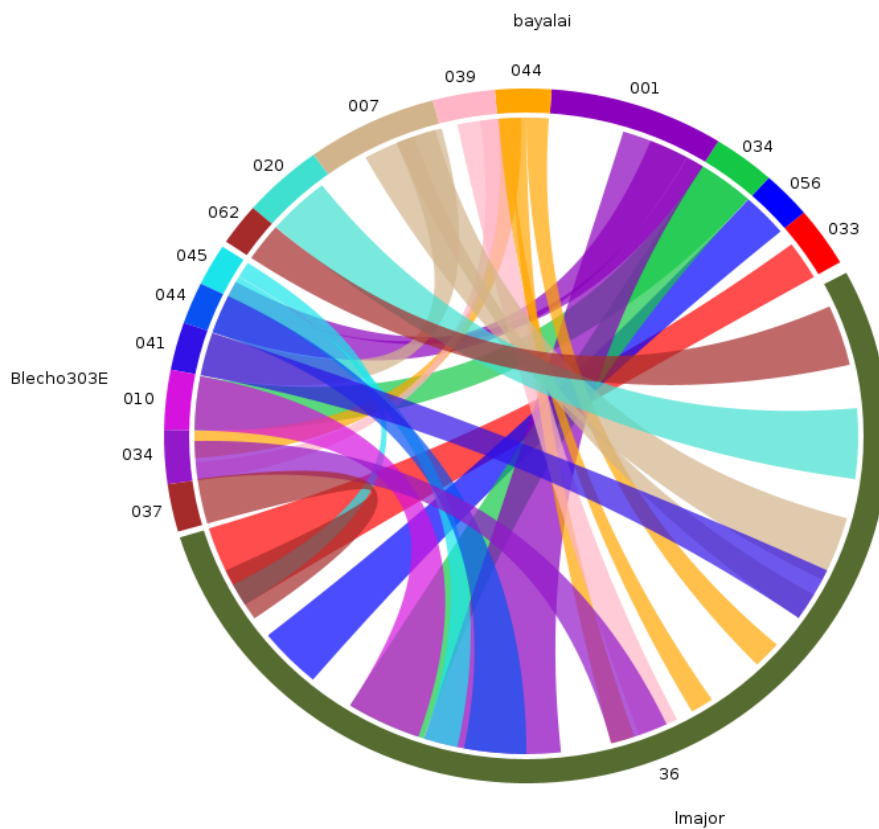


Figura 4: Gráfico da sintenia entre as espécies *Blechnomonas* sp. TCC303E, *B. ayalai* e *L. major*, gerada por SyMAP v5.0.6. O gráfico mostra os blocos sintênicos entre os scaffolds de *Blechnomonas* sp. TCC303E, os contigs de *B. ayalai* e o cromossomo 36 de *L. major*. Cada cor representa um bloco sintênico vinculado com laços de cores iguais aos blocos sintênicos das espécies *Blechnomonas* sp. TCC303E e *B. ayalai*.

Como observado na figura 4, a organização gênica e os comprimentos são consistentes com a literatura, que reporta uma elevada sintenia em grandes regiões de genes colineares, como observado entre tripanossomas e leishmânias, sendo a colinearidade implicada no controle da expressão temporal da expressão de certos genes (Flegontov et al. 2016; Ghedin et al. 2004; Ivens et al. 2005; Peacock et al. 2007).

V.5 Genômica comparativa de *Blechnomonas* sp. TCC303E baseada na classificação funcional

Com o fim de obter uma visão geral da fisiologia do nosso organismo e ver as similaridades e diferenças em relação aos outros organismos, nesta seção foram comparadas as distribuições das categorias funcionais KOGs das proteínas preditas de dez organismos usados no estudo (uma espécie

por gênero), as quais foram obtidas através do processo de anotação funcional pelo programa eggNOG. O número de anotações e as porcentagens que eles representam dos seus proteomas utilizados provêm dos seguintes organismos: *B. ayalai* B08-37 (4.511 anotações, 56,1%), *C. fasciculata* CfCl (5.699 anotações, 60,1%), *E. monterogei* LV88 (5.081 anotações, 61,3%), *L. major* Friedlin (5.294 anotações, 62,1%), *L. pyrrhocoris* H10 (5.733 anotações, 58%), *P. confusum* CUL13 (4.635 anotações, 53,5%), *Phytomonas* AKH_PRJEB1535 (3.627 anotações, 56,8%), *T. brucei* gambiense DAL972 (4.594 anotações, 57,5%), *B. saltans* LakeKonstanz (6.582 anotações, 36,1%). No caso de *Blechnomonas* sp. TCC303E, foram anotadas 4.604 proteínas, que representam 57,2% do proteoma total, restando 3.440 (42,8%) proteínas que não apresentam nenhuma categoria funcional KOG.

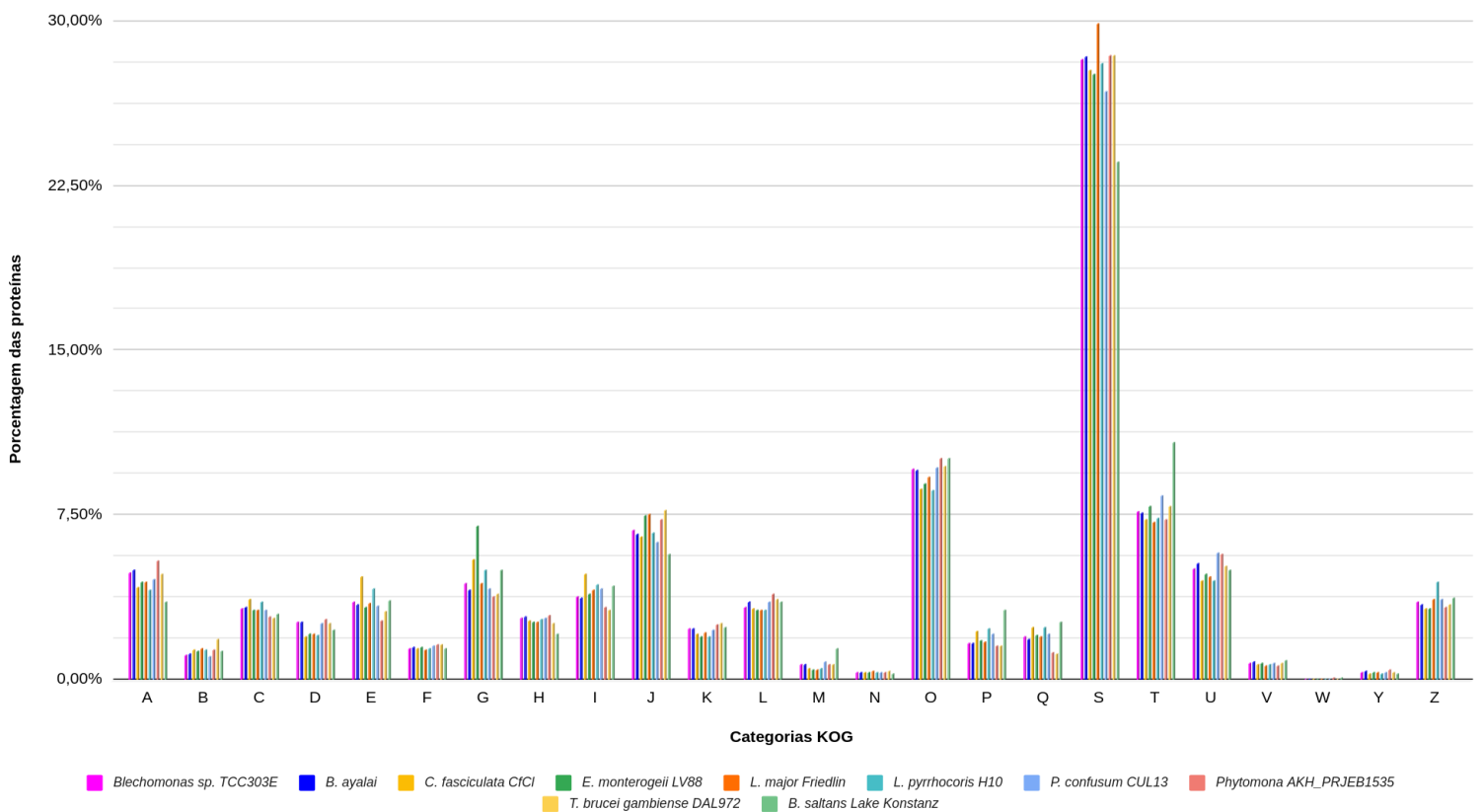


Figura 5. Comparação das quantidades dos proteomas de 9 tripanossomatídeos e de *B. saltans* nas categorias funcionais KOG geradas pela ferramenta eggNOG. No eixo horizontal estão as letras das categorias funcionais (24) e no vertical, a porcentagem de proteínas pertencentes a cada categoria funcional. Não foi considerada a categoria “R” nesta análise, pois nessa categoria estão as proteínas com função desconhecida (proteínas órfãs) e sem anotação pelo EggNOG.

Na figura 5, as distribuições das categorias funcionais de nosso organismo mostram tendências parecidas com o resto das espécies. Porém, se observam variações entre táxons, em alguns

casos de forma notória. Isso devido a os cinetoplastídeos compartilharem genes ancestrais e vias metabólicas, mas também na sua história evolutiva experimentaram contrações e expansões (inclusive intra espécie), incluindo genes específicos de linhagem (inovações), o que resultou em distribuições assimétricas (Jackson et al. 2016). Nesse contexto, também se poderia dizer que os genomas dos parasitas são simplificados se comparados com outros cinetoplastídeos de vida livre e eucariotos, embora tenham mantido e evoluído genes que permitem a adaptação às diferentes novas condições ambientais em seus hospedeiros. Por exemplo, genes que possibilitam as mudanças morfológicas durante os ciclos de vida, aproveitamento de nutrientes e interação com os hospedeiros (Jackson et al. 2016). Assim, entre as categorias com maiores quantidades de proteínas, estão as categorias envolvidas nas mudanças morfológicas e fisiológicas (O, J e A), além do aproveitamento de nutrientes (E, G, I e U) e na interação com o ambiente e hospedeiro (T).

Pode-se observar que a categoria com maior porcentagem de proteínas entre os táxons é a categoria “S” com quase 30%, e que representam às proteínas com funções ainda a serem descobertas ou rotuladas, as quais são chamadas de “proteínas hipotéticas”. Essas porcentagens destacadas na categoria S são geralmente observadas nos proteomas anotados dos tripanossomatídeos, como mencionado anteriormente, porcentagens essas que podem superar a metade do total dos proteomas (Callejas-Hernández et al. 2018; Silber e Pereira 2012). Assim nesta análise as espécies *L. major* Friedlin (29.88%) e *B. saltan* (23.57%) têm o maior e o menor número de proteínas hipotéticas, e as espécies do gênero *Blechnomonas* são quase idênticas, tendo *B. ayalai* 28.39% e *Blechnomonas* sp. TCC303E 28.24%.

A categoria “O” é a segunda mais abundante e representa proteínas relacionadas a funções em modificação pós-traducional, renovação de proteínas, e chaperonas. A abundância desta categoria nos cinetoplastídeos mostra uma alta dedicação para a renovação e controle metabólico. É conhecido que a adaptação e sobrevivência dos cinetoplastídeos a diferentes ambientes depende de uma regulação ajustada da expressão dos genes, e que acontece principalmente após a transcrição, nos processos de maturação do RNA, tradução e degradação (Clayton 2019). Modificações pós tradução observadas em histonas e num grande número de enzimas metabólicas, assim como em proteínas de união a RNA mensageiros em tripanossomatídeos, tiveram impacto no metabolismo e como consequência na patogênese dos parasitas. Além disso, essas modificações têm influência na resistência a drogas como foi observado para as proteínas HSP90 (Kalesh et al. 2021; Michels et al. 2021; Picchi et al. 2017; Zhang et al. 2020). Assim, no gráfico se observa que as menores quantidades de proteínas nesta categoria pertencem às espécies da subfamília Leishmaniinae, e as espécies *B. saltans* e *Phytomonas* têm as maiores quantidades. Essas proteínas apresentam funções como de ubiquitinação, hidrolases, peptidases, metilases e glicosilases. Ademais, *Blechnomonas* apresentou diverso número de cópias de famílias de chaperonas como as de choque térmico HSP90, HSP70, HSP60, HSP20, e de DNAJ. Também foram encontradas proteínas com atividades em tioredoxina,

glutationa peroxidase, triparedoxina e glutaredoxina.

A próxima categoria com maior número de anotações foi a categoria T, onde quantidades similares de proteínas classificadas nessa categoria foram observadas em outros tripanossomatídeos (Grisard et al. 2010; Tschoeke et al. 2014). Esta categoria abrange proteínas envolvidas nos mecanismos de transdução de sinais, mecanismos que são essenciais para a sobrevivência, adaptação e transmissão dos parasitas, pois informa as mudanças ambientais e coordena os processos intracelulares segundo as condições que experimentam. Porém, na figura 5 observa-se que a abundância dos parasitas é menor se comparada com a do organismo de vida livre *B. saltans* (com até 3,61% de diferença). Talvez essa diferença seja consequência de uma menor quantidade de estímulos dos ambientes que habitam esses parasitas (dentro de hospedeiros), mudança que também foi observado inclusive entre parasitas com diferentes preferências de habitats, por exemplo, em menor quantidade em parasitas intracelulares (*T. cruzi* e *L. major*) se comparado com o parasita extracelular *T. brucei*, possivelmente para evadir o sistema imune do hospedeiro alterando continuamente a expressão dos genes ESAG (Tagoe et al. 2015).

As características próprias dos mecanismos de tradução de sinais também contribuíram para diferenciar os cinetoplastídeos dos metazoários. Por exemplo, em tripanossomatídeos, várias classes de proteínas estão ausentes e outras se expandiram, como os membros das subfamílias de proteínas quinases CMGC e STE, o que sugere uma complexidade regulatória focada na proliferação, diferenciação, estresse e no ciclo celular. Os tripanossomatídeos apresentam um amplo e complexo repertório de proteínas quinases, onde a fosforilação da tirosina tem uma função importante na regulação do metabolismo –apesar de não terem tirosinas quinases conhecidas, contam com uma proteína quinase de dupla especificidade que pode fosforilar essa tirosina (Hammarton et al. 2003; Naula et al. 2005).

Nos tripanossomatídeos, a transdução de sinais é principalmente dependente de cAMP por meio da AC (adenilato ciclase), a qual é diferente estruturalmente das dos mamíferos. Além disso, não codificam proteínas G e no caso de *T. brucei* o PKA (*Protein Kinase A*) é inativo, e mais do que 70% do repertório das pequenas GTPases estão dedicadas a processos de transporte intracelular, secreção, endocitose e patogênese, o que mostra um papel diminuído das GTPases nos mecanismos de transdução de sinais, contrário e inusual comparado com o resto dos metazoários, mas com um sistema de tráfego relativamente complexo (Field 2005). Assim, em *Blechnomonas* sp TCC303E foram encontradas 36 GTPases e, destas, a maioria (19) é rotulada com a categoria U (dedicada a tráfego intracelular, secreção e transporte vesicular) e 7 com a categoria T, e no caso de *B. ayalai*, das 40 GTPases 21 da categoria U e 8 da categoria T. Nesta categoria também foram encontradas em *Blechnomonas* TCC303E algumas anotações de adenilato ciclase (0 em *B. ayalai*, talvez por problemas de montagem ou anotação), 117 proteínas com atividades quinase (128 em *B. ayalai*), e 4

fosfodiesterases (5 em *B. ayalai*). Além disso, foram encontrados membros da família CMGC e uma proteína quinase de dupla especificidade em *Blechnomonas*. Embora, esses repertórios de *Blechnomonas* sejam menores do que os repertórios dos tritryps, os mecanismos para perceberem os ambientes poderiam ser semelhantes aos hipotetizados para o resto de tripanossomatídeos (Beavo 1995; Field 2005; Kaiser 2019; Vij et al. 2014).

Outra categoria com alto número de proteínas é a categoria J, que junto à categoria O mantém uma elevada taxa de câmbio e controle metabólico nos cinetoplastídeos. A categoria J representa as proteínas envolvidas na tradução, estrutura ribossomal e biogênese. Na distribuição, as espécies *B. saltans* e *T. brucei* têm o menor e o maior número de proteínas, respectivamente, e entre tripanossomatídeos se podem detectar variações consideráveis. Estas variações são consequência do variável número de duplicações do material genético entre os tripanossomatídeos, eventos que permitiram a expansão de famílias de proteínas que integram diferentes categorias funcionais, sendo as famílias de proteínas com maior expansão as com função na superfície celular e as proteínas ribossomais. Certos genes podem ter mais do que uma função e podem estar ubicados em complexos moleculares distintos, como por exemplo as proteínas ribossomais, as quais são iniciadores da tradução e também participam na alongação (Naora 1999). Outras funções das proteínas ribossomais encontram-se na resistência contra drogas, como reportado para a subunidade maior 60S (60SRL23a) em *L. donovani* (Cordeiro-Da-Silva et al. 2001; Das et al. 2013; Hendrickx et al. 2021), e também contribuem para a disfunção do sistema imunológico do hospedeiro, como reportado para a proteína ribossomal S3a estudada em *L. major* (Cordeiro-Da-Silva et al. 2001; Das et al. 2013)

Na categoria G, envolvida no transporte e metabolismo de carboidratos, se pode observar uma distribuição peculiar, onde a espécie *E. monterogeei* se destaca com uma maior diferença no número de anotações seguido em abundância pelas espécies *C. fasciculata*, *L. pyrrocoris* e *B. saltans*. Apesar de os tripanossomatídeos terem uma similar potencial rede metabólica, eles apresentam variações importantes em conteúdo, função, estrutura e nos padrões de expansão das enzimas, diferenças originadas da longa divergência há mais de 100 milhões de anos (Michels et al. 2021). Por exemplo, *T. brucei* e *Phytomonas* evoluíram para aproveitar ambientes abundantes em glucose, dando prioridade a mecanismos de transporte e metabolismo dessas moléculas, porém esses organismos são os com menor número de anotações nesta categoria. A respeito disso, se faz necessário considerar os processos evolutivos particulares dos parasitas, que envolvem, por exemplo, os hábitos de alimentação dos vetores, os lugares de maior permanência ou preferência dos parasitas durante o ciclo de vida. Assim, entre os tritryps novos caminhos e formas de controle foram descobertas, sendo algumas específicas da espécie. Por exemplo, *T. brucei* apresenta menor quantidade de sugar quinases, provavelmente devido a uma restrita variedade de açúcares presentes no plasma na dieta do vetor, contrário aos vetores de *Leishmania* que além de sangue também consome néctares ricos em açúcares.

Além disso, o único que hidrolisa dissacarídeos é *L. major*, e *T. cruzi* produz transportadores para hexoses fosfatadas, as quais estão disponíveis no citoplasma (Berriman et al. 2005). Ademais, é conhecida a perda paralela de muitos genes metabólicos dos tripanossomas africanos e *Phytomonas* spp, especialmente esse último com o dobro de genes perdidos na evolução (Maslov et al. 2019). Outro aspecto a destacar do metabolismo dos carboidratos é sua influência na virulência dos tripanossomatídeos. Como exemplo, temos o manogênio (*mannogen*), oligossacarídeo produzido em um ramo específico da glicólise em *Leishmania* spp. e que pode atingir concentrações altas em estágios patogênicos dos parasitas, então pode servir como fator de virulência, além de armazenamento de carbono ou energia (Michels et al. 2021).

No caso da categoria U, que representa o tráfego intracelular, secreção e transporte vesicular, as espécies de ramos mais ancestrais (*B. saltans*, *P. confusum*, tripanossomas, *Phytomonas* e *Blechnomonas*) têm o maior número de proteínas anotadas, e no caso das espécies da subfamília Leishmaniinae têm a menor distribuição. O tráfego intracelular tem um papel crucial na progressão do ciclo celular, na infectividade e na evasão imune dos patógenos. Entre as proteínas com maior abundância estão as proteínas da família Rab, associadas com os processos de endocitose e exocitose, dessas se observa um grupo de genes conservados entre os tripanossomatídeos, e, ademais, alguns genes próprios para cada espécie. Por exemplo, Field et al. 2007 reportou as seguintes proteínas: Rab 23/28 (em tritryps), Rab14 (*T. cruzi* e *Leishmania*) e Rab32 (*T. cruzi*). Os repertórios de proteínas Rab entre as espécies do gênero *Blechnomonas* são similares e estão compostos por famílias conservadas entre eucariotos (Natesan et al. 2009): rab1, rab-2a, rab-4, rab-5, proteína rab6-like, rab7, rab11, rab-14, e rab18; no total, 15 proteínas em *Blechnomonas* sp. TCC303E e 14 em *B. ayalai*. A família Rab, junto à família de proteínas SNARE, contribui para um sistema relativamente complexo de tráfego intracelular nos tripanossomatídeos, se os compararmos com outros eucariotos como *S. cerevisiae*. Isso evidencia que em tripanossomatídeos essa maquinaria é conservada, porém pode apresentar diferenças particulares de cada espécie, como produto de caminhos evolutivos distintos, e o grau de divergência entre essas espécies (Field et al. 2007); Sem dúvidas, não são mais elaborados, se o compararmos com *B. saltans*, sendo essa simplicidade associada a uma diversidade fisiológica das eucariotos em vez de considerá-la como característica que deu origem ao parasitismo (Jackson et al. 2016).

Outras categorias com marcada diferença entre os tripanossomatídeos e *B. saltans* são as categorias A e H, a primeira se ocupa do processamento de RNA, enquanto a segunda categoria se ocupa da assimilação de coenzimas. As diferenças na categoria A talvez sejam consequência ou estejam relacionadas à alta taxa de substituição e controle metabólico dos parasitas, devido a seus ciclos de vida complexos (como mencionado anteriormente), e no caso da categoria H, a diferença poderia ser devida ao fato de que *B. saltans* adquire as coenzimas a partir das bactérias que ingere, ao contrário que nos tripanossomatídeos, que as obtêm a partir de seus hospedeiros. Por outro lado, nas

categorias M (biogênese da parede celular/membrana/envelope) e P (transporte e metabolismo de íons inorgânicos), se destaca a quantidade de proteínas no organismo de vida livre *B. saltans*. Essa redução de proteínas nestas categorias dos tripanossomatídeos tem consistência com a perda de redundâncias funcionais que experimentaram, como é o caso de algumas famílias de proteínas de membrana e transportadores iônicos (Jackson et al. 2016). Contudo, os dados refletem os eventos de redução de redundância funcional e diferenças nas distribuições de caracteres ancestrais entre os tripanossomatídeos, especialmente aqueles de famílias multigênicas, como as peptidases secretadas, proteínas de membrana e também elementos móveis (transposons).

V.6 Comparação genômica baseada nos domínios PFAM

Outro passo necessário na anotação funcional dos genes anteriormente mencionados são as anotações de seus domínios proteicos. Os domínios são regiões das proteínas que podem cumprir diversas funções, como estruturais, catalíticas e evolutivas, com as quais é possível inferir as funções de uma proteína, mesmo que às vezes não necessariamente seja uma simples soma de seus domínios. Nesta seção comparamos o número de cópias de 21 domínios, caracterizados em estudos prévios dos *Trypanosoma*, que foram encontrados nos proteomas de *Blechnomonas* sp. TCC303E, 10 outros tripanossomatídeos e o "bodonídeo" *B. saltans* (Berriman et al., 2005; Ivens et al, 2005; El-Sayed et al., 2005; Peacock et al., 2007). Estas anotações foram obtidas ao contrastar os proteomas com a base de dados PFAM (*Protein family*) utilizando o programa Hmmscan. Em *Blechnomonas* sp. TCC303E foram anotados 52,8% (4.249/8.044) de seu proteoma com 2.423 domínios PFAM únicos, onde 1.329 se apresentaram como cópias únicas e em 2,5% (204/8.044) foi anotado como domínios com função desconhecida o DUF (*Domain of Unknown Function*).

| | <i>Ble</i> | <i>Ba</i> | <i>Cf</i> | <i>Em</i> | <i>Ld</i> | <i>Lm</i> | <i>Lp</i> | <i>Ls</i> | <i>Tb</i> | <i>Ph</i> | <i>Pc</i> | <i>Bs</i> | Descrição | Anotação PF |
|---------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|--|-------------|
| PF00069 | 185 | 190 | 210 | 194 | 187 | 196 | 208 | 191 | 172 | 145 | 203 | 396 | Domínio de proteína-quinase | |
| PF00400 | 85 | 92 | 91 | 77 | 84 | 85 | 95 | 85 | 89 | 66 | 97 | 134 | Domínio WD, repetição G-beta | 396 |
| PF00076 | 65 | 68 | 80 | 70 | 69 | 72 | 78 | 61 | 64 | 53 | 61 | 68 | Reconhecimento ao RNA | 210 |
| PF00004 | 61 | 59 | 60 | 66 | 62 | 67 | 67 | 65 | 65 | 54 | 64 | 63 | Família ATPase (AAA) | 145 |
| PF00515 | 44 | 46 | 43 | 38 | 42 | 42 | 41 | 45 | 45 | 42 | 50 | 59 | Repetição de tetratricopeptídeo | 110 |
| PF13855 | 38 | 39 | 105 | 110 | 47 | 75 | 100 | 59 | 47 | 28 | 43 | 394 | Repetições ricas em leucina (LRR) | 61 |
| PF00005 | 38 | 41 | 55 | 50 | 41 | 46 | 62 | 40 | 22 | 27 | 40 | 77 | Transportador ABC | 36 |
| PF00153 | 29 | 30 | 36 | 34 | 33 | 39 | 36 | 31 | 28 | 25 | 34 | 38 | Transportador mitocondrial | < 11 |
| PF12796 | 28 | 27 | 33 | 26 | 27 | 29 | 30 | 33 | 25 | 16 | 29 | 105 | Repetições de Anquirina | |
| PF13920 | 23 | 21 | 28 | 15 | 22 | 25 | 29 | 25 | 22 | 18 | 27 | 40 | Dedo de zinco (C3HC4) | |
| PF01490 | 21 | 21 | 38 | 26 | 23 | 28 | 39 | 28 | 33 | 12 | 19 | 16 | Transportador transmembrana de aminoácidos | |
| PF02493 | 18 | 20 | 22 | 21 | 21 | 21 | 22 | 21 | 15 | 15 | 22 | 29 | Repetição MORN | |
| PF00173 | 18 | 21 | 26 | 22 | 20 | 22 | 24 | 20 | 9 | 10 | 16 | 19 | União ao Heme/esteróide | |
| PF01457 | 17 | 13 | 18 | 176 | 3 | 6 | 8 | 8 | 6 | 3 | 9 | 14 | Leishmanolisina | |
| PF07344 | 13 | 17 | 51 | 18 | 28 | 61 | 36 | 26 | 5 | 6 | 16 | 12 | Glicoproteína de Superfície Amastina | |
| PF00642 | 13 | 7 | 16 | 11 | 11 | 12 | 15 | 16 | 11 | 8 | 9 | 15 | Dedo de zinco (C-x8-C-x5-C-x3-H) | |
| PF00806 | 11 | 11 | 9 | 10 | 9 | 10 | 11 | 11 | 10 | 9 | 12 | 13 | Família Pumilio de União ao RNA | |
| PF00012 | 11 | 12 | 14 | 19 | 15 | 16 | 16 | 12 | 14 | 9 | 11 | 18 | Proteína Hsp70 | |
| PF03133 | 10 | 10 | 9 | 9 | 9 | 9 | 9 | 9 | 10 | 8 | 9 | 15 | Família ligase tubulina-tirosina | |
| PF00612 | 9 | 9 | 6 | 12 | 7 | 9 | 7 | 8 | 9 | 7 | 10 | 22 | União à calmodulina | |
| PF01187 | 1 | 0 | 1 | 1 | 2 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | Inibição da migração do macrófagos (MIF) | |

Figura 6. Comparação dos domínios proteicos anotados com domínios PFAM de 11 espécies de tripanossomatídeos e de *B. saltans*: *Blechnomonas* sp. TCC303E (*Ble*), *B. ayalai* (*Ba*), *C. fasciculata* (*Cf*), *E. monterogeei* (*Em*), *L. donovani* (*Ld*), *L. major* (*Lm*), *L. pyrrocoris* (*Lp*), *L. seymouri* (*Ls*), *T. brucei* (*Tb*), *Phytomonas* AKH_PRJEB1535 (*Ph*), *P. confusum* (*Pc*), *B. saltans* (*Bs*). O heatmap mostra nas colunas as

espécies e nas linhas os códigos PFAM relacionados com seus domínios putativos. Quantidades maiores de domínios são marcadas com cor azul mais escura.

O gráfico heatmap (figura 6) mostra o número de cópias dos domínios, organizados de forma decrescente em relação ao número de cópias no *Blechnomonas* sp. TCC303E, onde o maior número de cópias recebe a cor azul mais escura. Do gráfico se observa que as espécies de *Blechnomonas* têm similar número de cópias, com os seguintes 5 domínios como os mais abundantes, em ordem decrescente: PF00069, PF00400, PF00076, PF00004 e PF00515. Em termos gerais, todos os organismos testados mostram valores próximos, com algumas exceções notórias nos domínios PF13855 e PF01457. Outrossim, *B. saltans* se destaca do resto de espécies de forma clara com maiores números de cópias na grande maioria de domínios, o que era de se esperar segundo a literatura (Jackson et al. 2016). Estas diferenças relativas entre tripanossomatídeos e *B. saltans* coincidem com as anotações embasadas em categorias KOG do capítulo anterior, e são consistentes com os eventos evolutivos de redução de funções redundantes em tripanossomatídeos em vários domínios envolvidos em uma ampla variedade de processos biológicos.

O domínio PF13855 representa repetições em tandem, usualmente ricas em leucina (LRR, *Leucine-rich repeats*). LRR é amplamente conservado, desde vírus a eucariotos, e devido a constituir um módulo “comum” para construir estruturas de interação, oferece diversas funções para as proteínas, como interação proteína-proteína, reconhecimento de patógenos e transdução de sinais. Ademais, estudos mostram a participação do domínio LRR como elemento chave para a função de proteínas utilizadas por parasitas no processo de invasão a hospedeiros, como foi observado em um primeiro momento em bactérias. Depois, em tripanossomatídeos, se observou o domínio LRR formando parte de proteínas de superfície celular altamente imunogênicas, como PPG e PSA-2, que são utilizadas por *Leishmania* para invadir macrófagos (Kedzierski et al. 2004) e por outros parasitas para interagir com seus hospedeiros, como *Plasmodium falciparum* (Pierrot et al. 2012). O domínio PF13855 foi encontrado em 0,47% (38/8.044) do proteoma de *Blechnomonas* sp. TCC303E, porcentagem composta principalmente por proteínas anotadas como hipotéticas e o resto envolvidas nos processos de transdução de sinais, modificação e mudanças de proteínas, e mecanismos de defesa. Por sua parte, os tripanossomatídeos que destacam com maiores quantidades são *E. monterogeii* (724), *C. fasciculata* (534), *L. pyrrhocoris* (480) e *L. major* (390), e no caso de *B. saltans* (2470), este domínio é maior que 3,4 vezes a maior quantidade entre os tripanossomatídeos (*E. monterogeii* com 724).

O domínio PF00400 caracteriza a superfamília de proteínas WDRs (*WD40 repeats*), uma das mais abundantes e com grande diversidade de funções nos eucariotos que se originou nas etapas iniciais da separação com os procariotos. O domínio WD40 permite interações proteína-proteína e junto a outros domínios originam complexas estruturas proteicas com ampla variedade de funções

(Campbell et al. 2004; van Nocker e Ludwig 2003). Esta versatilidade exibida por esta família de proteínas derivou em proteínas altamente conectadas e centralizadas nas redes celulares, participando de múltiplos processos celulares como: processamento de RNA, transdução de sinais, tráfico vesicular, regulação da divisão celular, apoptose, ubiquitinação/degradação de proteínas, montagem da cromatina, remodelação e regulação transcricional (Chahar et al. 2015; Li et al. 2014). Em tripanossomatídeos, membros da família WDR, como RACK1, participam dos processos de sinalização celular, divisão, diferenciação e tradução. Ortólogos de RACK1 são essenciais para a virulência e patogenicidade em *Leishmania* e *Trypanosoma*. Em *Leishmania*, LACK, é essencial para a viabilidade do parasita, a sobrevivência na temperatura do hospedeiro e a infecção robusta dos macrófagos do hospedeiro (Choudhury et al. 2011). Em TCC303E foram encontrados 146 domínios WD40, e 143 em *B. ayalai*, o que representa 1,82% (146/8.044) e 1,78% (143/8.042), respectivamente. A maior parte delas são proteínas hipotéticas, seguidas de proteínas para amadurecimento do RNA, transporte intracelular, citoesqueleto, controle do ciclo celular, transcrição, transdução de sinais e modificação de cromatina.

O domínio PF00069 é o mais abundante e representa os domínios quinase, altamente conservados nos tripanossomatídeos e integra uma considerável porcentagem de seus proteomas (ao redor de 2% em *Trityps*); e, por sua vez, estes domínios apresentam algumas divergências específicas de sequências próprias de cada espécie (Merritt et al. 2014; Tschoeke et al. 2014). As quinases têm a função de fosforilar proteínas, mecanismo chave para a regulação de uma extensa variedade de processos celulares fundamentais para o parasita, como foi mencionado na análise com os termos KOG, formando parte dos processos de transdução de sinais. Em tripanossomatídeos, as quinases envolvidas na transdução de sinais terminam na regulação da expressão dos genes em nível pós-transcricional, já que ao contrário que em outros eucariotos, que regulam no nível transcricional, os tripanossomatídeos carecem de promotores de transcrição e a transcrição é feita em blocos policistrônicos (Parsons et al. 2005; Wiese et al. 2003). Ademais, as quinases são chave na interação hospedeiro-parasita (Silvestre et al. 2022). Por exemplo, *Leishmania* utiliza adenilato quinase para reduzir o ATP externo e dessa forma prevenir a apoptose do macrófago, assegurando sua sobrevivência (Kulkarni et al. 2019). Este domínio é mais numeroso no gênero *Blechnomonas* e representa 2,3% (185/8.044) do proteoma de *Blechnomonas* sp. TCC303E (2,4% em *B. ayalai*). A maior parte destas proteínas está envolvida na transdução de sinais, seguido por proteínas que controlam o ciclo celular, proteínas hipotéticas, transporte e metabolismo de carboidratos e coenzimas, transcrição, tradução, reparação, recombinação e replicação do DNA, e 9 genes contendo domínios de adenilato quinase nas espécies *Blechnomonas*.

O domínio PF00076 é uma das classes de domínio mais comuns em eucariotos e em alguns tripanossomatídeos, incluindo *Blechnomonas*, e o terceiro em abundância no heatmap. Este domínio capacita as proteínas a unir-se a transcritos de RNA (RBPs, *RNA-binding protein*) mediante seu

dedicado domínio funcional RBD (*RNA-binding domain*; também conhecido como RRM, *RNA recognition motif*). O tipo de classe das RBPs depende do motivo que constitui o RBD, como por exemplo os domínios RRM, dedos de zinco, Pumilio e ALBA, para nomear alguns dos mais comuns (Romagnoli et al. 2020; Romaniuk et al. 2016). Os RBPs constituem ao redor de 3-10% do genoma de eucariotos (Glisovic et al. 2008). Se somarmos os domínios RBD com outros domínios de união ao RNA (como dedos de zinco e Pumilio, junto a outros domínios não mostrados no heatmap), resulta em uma elevada quantidade, consistente com sua grande importância para a função celular dos tripanossomatídeos, já que regulam a expressão gênica no nível pós-transcricional ao unir-se aos RNA mensageiros, o que determina a abundância dos transcritos e a eficiência de suas traduções. Estas proteínas por sua vez são reguladas por enzimas como as quinases e as fosfatases, anteriormente mencionadas (Ivens et al. 2005; Romaniuk et al. 2016).

O domínio PF00004 faz referência à superfamília de proteínas chamadas AAA+ (*ATPases associated with various cellular activities*), que utilizam a energia liberada da hidrólise de ATP para gerar eventos de remodelação molecular (energia mecânica) que são empregados em uma ampla variedade de atividades celulares (Snider et al. 2008; Wendler et al. 2012). No proteoma de *Blechnomonas* sp. TCC303E, cerca de 0,8 % (61/8.044) das proteínas contêm o domínio AAA+, e se distribuem em diversas funções (similar em *B. ayalai*), principalmente nas modificações pós-transcricionais e na renovação das proteínas, como a proteína cdc48, central nos processos de controle de qualidade de proteínas para eliminar proteínas danificadas, o que protege a *Leishmania* contra os danos recebidos dentro do macrófago (Aguilar et al. 2020); também foram encontradas proteínas envolvidas em manutenção, replicação e recombinação do DNA, como a família de proteínas RuvB que participa na progressão do ciclo celular e remodelamento da cromatina (Ahmad e Tuteja 2012); além de proteínas com função no citoesqueleto, como as proteínas da cadeia pesada da dineína (DHC) e uma proteína katanin-like envolvida no controle do tamanho do flagelo em tripanossomatídeos (Casanova et al. 2009).

O domínio PF00515, chamado repetições dos tetratricopeptídeos ou TPR (*The tetratricopeptide repeats*), está entre os mais abundantes, como era de se esperar devido a apresentar uma grande diversidade de sequências que permitem interações proteína-proteína e proteína-lípidos, envolvidas na regulação do ciclo celular, função de chaperona e modificações pós-traducionais; Ainda, sua estrutura serve como andaime para a formação de diferentes complexos multiproteicos, receptor de importação peroxissomal e NADPH oxidase (Diniz et al. 2010). Foram encontrados em aproximadamente 0,55% (44/8.044) do proteoma de *Blechnomonas* sp. TCC303E e em *B. ayalai* em 0,57% (46/8.042) de seu proteoma. Estas proteínas estão envolvidas em funções celulares como: Estrutura da cromatina, controle do ciclo celular, chaperonas, modificações pós-traducionais,

renovação das proteínas, transporte de carboidratos e substâncias inorgânicas, transdução de sinais e citoesqueleto.

O domínio PF01457 caracteriza a família GP63, contabilizadas como 0,21% (17/8.044) do proteoma de *Blechnomonas* sp. TCC303E e 0,16% em *B. ayalai* (13/8.042) de seu proteoma e de forma notória em 2,12% (176/8.285) do proteoma de *E. monterogeei*. Este último tripanossomatídeo tem uma inusual etapa intraeritrocitária, e de forma distinta de *Leishmania*, as sequências de GP63 são funcionalmente inativas (como metaloproteinases) na etapa promastigota, pelo que se especula que tenha uma expressão específica intraeritrocitária (Medina-Acosta et al. 1994; Yao 2010). Talvez estas diferenças tenham relação com um repertório maior de GP63 para conseguir parasitar esse particular ambiente intraeritrocitário.

Nos tripanossomatídeos, uma das moléculas chave para a interação hospedeiro-parasita é a glicoproteína GP63, também chamada de MSP (*major surface protease*), metaloprotease zinc, protease ácida de superfície, protease de superfície de promastigota, e leishmanolisina, a qual se apresenta maiormente unida a superfície celular por meio do ancoragem GPI (*glycosylphosphatidylinositol*), porém, também é liberada ao meio em forma livre ou dentro de vesículas. Além disso, as GP63 são uma família multigênica que pode variar de forma considerável em número de cópias dependendo da espécie (Castro Neto et al. 2019).

Estas proteínas apresentam diferenças em suas sequências, no momento em que são expressadas e em sua localização, o que resultou na expressão de distintas funções dependendo da espécie, inclusive dentro de um mesmo parasita estas sequências podem manifestar funções colaborativas ou redundantes; estas funções são importantes para os processos de infecção e sobrevivência em seus hospedeiros invertebrados e vertebrados e pode ser que tenham facilitado a adaptação a novos ambientes (Yao 2010). Por exemplo, em *Leishmania*, é crucial para sua patogenicidade, evadir o sistema imune de seu hospedeiro, interagir com a matriz extracelular e promover uma efetiva fagocitose pelos macrófagos; em *T. brucei*, ajuda na renovação da capa de VSG; em *T. cruzi*, se observou in vitro que a GP63 é necessária para invadir as células mamíferas (Llewellyn et al. 2015). No entanto, pouco se conhece da função que desempenha a GP63 em invertebrados, porém se hipotetiza que sua função primária ou ancestral estava relacionada com a sobrevivência em dípteros (Santos et al. 2006), o que derivou em possíveis funções na defesa, nutrição e aderência ao intestino de insetos. Sem dúvidas, estudos demonstraram a capacidade dos monoxenos de invadir e sobreviver em células fagocíticas e no fagocíticas de hospedeiros mamíferos, onde a GP63 teve uma função crucial nesses processos, o que corrobora os reportes de infecções oportunistas dos monoxenos, incluindo *B. pulexsimulantis* (d'Avila-Levy et al. 2014).

Outra proteína importante na interação hospedeiro-parasita, caracterizada pelo domínio PF07344, é a glicoproteína amastina, que se classifica em quatro clases (α , β , γ , and δ). A amastina

promove a fagocitose de *Leishmania* ao interagir com a membrana dos macrófagos e viabilizar sua sobrevivência intracelular em sua forma amastigota (de Paiva et al. 2015). Em nosso organismo foram encontradas 13 (0,16%) proteínas com o domínio PF07344 e 17 (0,21%) em *B. ayalai*, e de forma notória, à exceção de *E. monterogeei*, no grupo de Leishmaniinae o número de amastinas é maior. O caso de *E. monterogeei* talvez seja explicado pelas funções das proteínas multigênicas que se sobrepõem de forma sinérgica ou redundante, assim se um falhar o outro o compensa, como no caso dos genes MSP e LPG em *Leishmania* (Yao 2010).

As quantidades de domínios tem correlação em abundância de repertório funcional com a análise anterior de termos KOG que expõem processos evolutivos de contrações e expansões de características ancestrais, principalmente de famílias multigênicas de superfície e peptidases (El-Sayed et al. 2005b; Jackson 2015; Jackson et al. 2016; Tschoeke et al. 2014). Ademais, as principais funções moleculares dos domínios descritas nestas análises foram as de união ou interação de proteínas centralizadas (múltiplas conexões e funções) com proteínas e nucleotídeos, focados principalmente na regulação pós-transcricional, na tradução e pós-tradução; e também focadas na interação com receptores e transdutores de sinal. Sendo essas descrições consistentes com a abundância de termos GO descritos na literatura (Tschoeke et al. 2014).

V.7 Análise das famílias de genes usando a abordagem de clustering

Nesta etapa se classificou o proteoma de *Blechnomonas* sp. TCC303E junto ao de 11 espécies de tripanossomatídeos em grupos ortólogos (GOs) por meio do método de clustering ou agrupamento por relações de ortologia, para o qual se utilizou a ferramenta OrthoFinder (Tabela 5). Os proteomas somaram o total de 110.009 proteínas, das quais 101.008 (~91,8%) foram agrupadas em 10.753 GOs (contendo ao menos 2 proteínas), com 2803 GOs contendo proteínas de cópia única e 6992 (6,4%) proteínas específicas de espécies agrupadas em 1497 GOs. O G50 estimado foi de 12 (50% dos genes estão em GOs desse tamanho ou maior). Do resto das 9.001 (8,2%) proteínas classificadas como sequências órfãs ou singleton, 341 proteínas são de *Blechnomonas* sp. TCC303E (4,2% de seu proteoma). O resto das proteínas de *Blechnomonas* sp. TCC303E (7703) formaram 7.159 GOs, número próximo ao de *B. ayalai* (7.295) e em um contexto onde *L. pyrrocoris* é a espécie com maior número (8.089) e *Phytomonas* tem o menor (5.497). Por sua vez, *Blechnomonas* sp. TCC303E apresenta 143 proteínas (1,8%) em 28 GOs específicos da espécie. *B. saltans* (5632) e o tripanossomatídeo mais próximo, *P. confusum* (184), contêm a maior quantidade de proteínas específicas da espécie, e as menores quantidades se encontram nas espécies *L. donovani* (8) e *L. seymouri* (11).

Tabela 5. Descrição das categorias dos resultados obtidos com o programa OrthoFinder para 11 espécies de tripanossomatídeos e *B. saltans*: *Blechnomonas* sp. TCC303E (*Ble*), *B. ayalai* (*Ba*), *C. fasciculata* (*Cf*), *E. monterogeei* (*Em*), *L. donovani* (*Ld*), *L. major* (*Lm*), *L. pyrrocoris* (*Lp*), *L. seymouri* (*Ls*), *T. brucei* (*Tb*), *Phytomonas* AKH_PRJEB1535 (*Ph*), *P. confusum* (*Pc*), *B. saltans* (*Bs*).

| Categorias/Espécies | <i>Ble</i> | <i>Ba</i> | <i>Cf</i> | <i>Em</i> | <i>Ld</i> | <i>Lm</i> | <i>Lp</i> | <i>Ls</i> | <i>Ph</i> | <i>Tb</i> | <i>Pc</i> | <i>Bs</i> |
|---|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Número de proteínas | 8044 | 8042 | 9489 | 8285 | 8023 | 8519 | 9878 | 8488 | 6381 | 7988 | 8659 | 18213 |
| Número de proteínas em GOs | 7703 | 7829 | 9352 | 8059 | 7875 | 8494 | 9588 | 8272 | 5761 | 7267 | 7533 | 13275 |
| Número de singletons | 341 | 213 | 137 | 226 | 148 | 25 | 290 | 216 | 620 | 721 | 1126 | 4938 |
| Porcentagem de proteínas em GOs | 95.8 | 97.4 | 98.6 | 97.3 | 98.2 | 99.7 | 97.1 | 97.5 | 90.3 | 91.0 | 87.0 | 72.9 |
| Porcentagem de singletons | 4.2 | 2.6 | 1.4 | 2.7 | 1.8 | 0.3 | 2.9 | 2.5 | 9.7 | 9.0 | 13.0 | 27.1 |
| Número de GOs contendo espécies | 7159 | 7295 | 7998 | 7098 | 7463 | 7549 | 8089 | 7836 | 5497 | 6379 | 6900 | 7342 |
| Porcentagem de GOs contendo espécies | 66.6 | 67.8 | 74.4 | 66.0 | 69.4 | 70.2 | 75.2 | 72.9 | 51.1 | 59.3 | 64.2 | 68.3 |
| Número de GOs específicos da espécie | 28 | 21 | 25 | 25 | 4 | 10 | 53 | 5 | 9 | 85 | 55 | 1177 |
| Número de proteínas em GOs específicos da espécie | 143 | 149 | 101 | 91 | 8 | 77 | 213 | 11 | 20 | 363 | 184 | 5632 |
| Porcentagem de proteínas em GOs espécie-específicos | 1.8 | 1.9 | 1.1 | 1.1 | 0.1 | 0.9 | 2.2 | 0.1 | 0.3 | 4.5 | 2.1 | 30.9 |

Em um enfoque de grupos de ortólogos por pares observados no heatmap (figura 7), a espécie *L. pyrrocoris* compartilha os maiores números de ortólogos com as espécies *C. fasciculata* (7.897) e *L. seymouri* (7.768). No outro extremo, se encontra a espécie *B. saltans*, dividindo as menores quantidades de ortólogos com os tripanossomatídeos, como era de se esperar devido a sua maior distância filogenética, onde o menor número compartilhado (4.355) é mostrado com *Phytomonas*. Por sua vez, *Blechnomonas* sp. TCC303E compartilha os maiores números de ortólogos com as espécies *B. ayalai* (7.013), *L. pyrrocoris* (6.511) e o resto do grupo das Leishmaniinae.

| Táxon | <i>Ble</i> | <i>Ba</i> | <i>Cf</i> | <i>Em</i> | <i>Ld</i> | <i>Lm</i> | <i>Lp</i> | <i>Ls</i> | <i>Ph</i> | <i>Tb</i> | <i>Pc</i> | <i>Bs</i> |
|------------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| <i>Ble</i> | 7159 | 7013 | 6494 | 6145 | 6339 | 6411 | 6511 | 6458 | 5159 | 5824 | 5946 | 5241 |
| <i>Ba</i> | 7013 | 7295 | 6646 | 6291 | 6469 | 6544 | 6665 | 6586 | 5239 | 5940 | 6033 | 5346 |
| <i>Cf</i> | 6494 | 6646 | 7998 | 6940 | 7278 | 7370 | 7897 | 7707 | 5324 | 6002 | 6224 | 5551 |
| <i>Em</i> | 6145 | 6291 | 6940 | 7098 | 6833 | 6908 | 6977 | 6888 | 5101 | 5722 | 5798 | 5197 |
| <i>Ld</i> | 6339 | 6469 | 7278 | 6833 | 7463 | 7434 | 7302 | 7202 | 5221 | 5890 | 6020 | 5362 |
| <i>Lm</i> | 6411 | 6544 | 7370 | 6908 | 7434 | 7549 | 7391 | 7284 | 5281 | 5967 | 6093 | 5426 |
| <i>Lp</i> | 6511 | 6665 | 7897 | 6977 | 7302 | 7391 | 8089 | 7768 | 5351 | 6018 | 6236 | 5564 |
| <i>Ls</i> | 6458 | 6586 | 7707 | 6888 | 7202 | 7284 | 7768 | 7836 | 5290 | 5949 | 6175 | 5481 |
| <i>Ph</i> | 5159 | 5239 | 5324 | 5101 | 5221 | 5281 | 5351 | 5290 | 5497 | 4928 | 4847 | 4355 |
| <i>Tb</i> | 5824 | 5940 | 6002 | 5722 | 5890 | 5967 | 6018 | 5949 | 4928 | 6379 | 5703 | 5094 |
| <i>Pc</i> | 5946 | 6033 | 6224 | 5798 | 6020 | 6093 | 6236 | 6175 | 4847 | 5703 | 6900 | 5669 |
| <i>Bs</i> | 5241 | 5346 | 5551 | 5197 | 5362 | 5426 | 5564 | 5481 | 4355 | 5094 | 5669 | 7342 |

Figura 7. Comparação do número de grupos ortólogos compartilhados pelas 11 espécies de tripanossomatídeos e *B. saltans* gerados pela ferramenta OrthoFinder: *Blechnomonas* sp. TCC303E (*Ble*), *B. ayalai* (*Ba*), *C. fasciculata* (*Cf*), *E. monterogeei* (*Em*), *L. donovani* (*Ld*), *L. major* (*Lm*), *L. pyrrocoris* (*Lp*), *L. seymouri* (*Ls*), *T. brucei* (*Tb*), *Phytomonas* AKH_PRJEB1535 (*Ph*), *P. confusum* (*Pc*), *B. saltans* (*Bs*).

Entre as proteínas específicas da espécie encontradas em *Blechnomonas* sp. TCC303E foram encontradas proteínas ribossômicas L37Ae, L1Tc, VSP, proteínas de união ao RNA, quinesinas e elementos transponíveis tipo TATE, além das proteínas hipotéticas. Elementos transponíveis (TEs, *transposable elements*), são sequências genômicas dinâmicas com o potencial de modificar a estrutura e perfil funcional do genoma. Nos parasitas que infectam mamíferos, os TEs influenciam na função e expressão dos genes; por exemplo, nos tripanossomatídeos, foi demonstrado que os TEs podem participar nos processos de transcrição, segregação dos cromossomos e no amadurecimento do RNA (Bringaud et al. 2008; Thomas et al. 2010), assim como também em eventos de recombinação que

podem derivar na amplificação genética (Pita et al. 2019).

Em *Leishmania* foi sugerido que os transposons tipo TATE (Elementos Transponíveis Associados ao Telômero) poderiam exercer um papel central na regulação e estrutura genética (Llanes et al. 2015; Tschoeke et al. 2014). Mecanismos que controlam os efeitos deletérios que podem causar a expansão dos TEs são necessários, como a maquinaria dos RNAs de interferência (iRNA) encontrados nos nematóides e em *T. brucei*. Em *Blechnomonas* sp. TCC303E não foi encontrado um mecanismo completo para iRNA, como observado em *T. cruzi* e *L. maior*, o que poderia sugerir mecanismos alternativos em certos tripanossomatídeos (El-Sayed et al. 2005b).

V.8 Expansões de família de genes dentro dos Clusters de Grupos Ortólogos

Com o propósito de observar a magnitude da expansão dos grupos ortólogos mais abundantes, foram escolhidos os 23 GOs mais abundantes, presentes em todos os organismos e com anotação funcional conhecida. Como mencionado acima, referindo-se ao trabalho de (Jackson et al. 2016), observa-se (figura 8) que os grupos ortólogos que apresentam maiores expansões são aqueles com funções relacionadas a mudanças morfológicas durante os ciclos biológicos, uso de nutrientes e interação com seus hospedeiros, características que conferem vantagens aos parasitas para se adaptarem às diferentes condições ambientais encontradas em seus hospedeiros. Destacam-se particularmente por apresentarem uma grande variedade de expansões os grupos ortólogos compostos pelo receptor adenilato ciclase (OG0000005) e a proteína leishmanolisina (OG0000006) em *T. brucei* e *E. monterogeii*, respectivamente; seguido pelos grupos ortólogos OG00000010 e OG00000011, que incluem transportadores ABC e proteínas E (Ullman 1995).

| OG | Ble | Ba | Cf | Em | Ld | Lm | Lp | Ls | Ph | Tb | Pc | Bs |
|---|-----|----|----|----|----|----|----|----|----|----|----|----|
| Cadeia pesada de dineína (OG0000003) | 11 | 12 | 10 | 11 | 11 | 11 | 11 | 11 | 11 | 10 | 10 | 15 |
| Receptor Adenilato ciclase tipo a (OG0000005) | 2 | 1 | 7 | 28 | 4 | 6 | 6 | 2 | 1 | 53 | 1 | 4 |
| Leishmanolisina (OG0000006) | 16 | 4 | 11 | 48 | 1 | 1 | 4 | 2 | 3 | 4 | 7 | 4 |
| Transportador ABC (OG0000010) | 7 | 9 | 10 | 7 | 8 | 9 | 8 | 8 | 1 | 1 | 9 | 6 |
| Proteína E (OG0000011) | 4 | 5 | 12 | 10 | 6 | 8 | 13 | 9 | 1 | 1 | 8 | 5 |
| Acil graxo-CoA sintetase 1 (OG0000019) | 5 | 5 | 9 | 6 | 5 | 5 | 7 | 3 | 3 | 5 | 5 | 3 |
| Proteína de ligação de RNA (OG0000021) | 3 | 4 | 7 | 7 | 7 | 7 | 5 | 4 | 2 | 3 | 7 | 2 |
| Fosfatase serina treonina (OG0000025) | 3 | 3 | 7 | 6 | 5 | 6 | 9 | 2 | 4 | 5 | 3 | 2 |
| Calpaina (OG0000027) | 4 | 6 | 3 | 10 | 3 | 3 | 4 | 6 | 2 | 5 | 8 | 1 |
| Transportador de aminoácidos (OG0000035) | 2 | 2 | 7 | 2 | 4 | 4 | 6 | 5 | 3 | 2 | 10 | 1 |
| Caseína quinase I (OG0000038) | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 3 | 4 | 4 |
| ATPase transportadora de cálcio (OG0000044) | 3 | 3 | 4 | 5 | 4 | 4 | 4 | 4 | 2 | 2 | 3 | 5 |
| Proteína relacionada a HSP70 (OG0000046) | 3 | 3 | 4 | 3 | 3 | 4 | 6 | 3 | 1 | 4 | 3 | 5 |
| Subunidade ATPase 9 (OG0000057) | 4 | 3 | 5 | 4 | 2 | 3 | 5 | 2 | 2 | 2 | 2 | 3 |
| O-aciltransferase (OG0000063) | 2 | 2 | 6 | 3 | 2 | 6 | 4 | 4 | 1 | 1 | 3 | 1 |
| Putativa 3'-nucleotidase/nuclease (OG0000068) | 2 | 1 | 3 | 2 | 3 | 9 | 3 | 3 | 1 | 3 | 2 | 2 |
| Quinase serina/treonina (OG0000076) | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 10 | 1 |
| Kinesina (OG0000087) | 2 | 2 | 3 | 4 | 3 | 3 | 3 | 4 | 2 | 2 | 2 | 1 |
| Transportadora de membrana (OG0000090) | 3 | 3 | 2 | 1 | 2 | 2 | 2 | 2 | 2 | 7 | 3 | 2 |
| Fumarato redutase (OG0000096) | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 1 | 3 |
| Proteína RING-box (OG0000108) | 3 | 1 | 3 | 2 | 2 | 2 | 3 | 1 | 3 | 3 | 3 | 3 |
| Proteína relacionado com a hemolisina-III (OG0000115) | 3 | 3 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 | 6 |
| Ecotina (OG0000116) | 4 | 4 | 2 | 3 | 2 | 2 | 2 | 2 | 2 | 1 | 3 | 1 |

Figura 8. *Heatmap* dos 23 grupos ortólogos mais abundantes com anotação conhecida e presentes em todas as 11 espécies de tripanossomatídeos e *B. saltans* analisadas com o programa OrthoFinder. Linhas mostram os grupos ortólogos com suas descrições e as colunas, as espécies: *Blechnomonas* sp. TCC303E (*Ble*), *B. ayalai* (*Ba*), *C. fasciculata* (*Cf*), *E. monterogeii* (*Em*), *L. donovani* (*Ld*), *L. major* (*Lm*), *L. pyrrocoris* (*Lp*), *L.*

seymouri (*Ls*), *T. brucei* (*Tb*), *Phytomonas* AKH_PRJEB1535 (*Ph*), *P. confusum* (*Pc*), *B. saltans* (*Bs*). Cores mais escuras indicam maior número de cópias gênicas por grupo ortólogo.

Entre os grupos ortólogos relacionados às alterações morfológicas encontramos o grupo OG0000019 que abriga a sintetase acil graxo-CoA, que participa da ativação de ácidos graxos para a formação de membranas celulares e estruturas de superfície celular como as GPIs, que servem como âncoras para várias proteínas de superfície celular, como as proteínas VSGs em *T. brucei*. Neste parasita, a sintetase acil graxo-CoA está localizada em arrays, em tandem e é expressa constitutivamente em ambas as fases do ciclo de vida (Jiang e Englund 2001; Smith e Bütikofer 2010). Observa-se que, com exceção de *L. seymouri* e *Phytomonas*, todos os parasitas apresentam um maior número de cópias do que *B. saltans*, destacando-se *C. fasciculata* com 9 cópias.

Também temos o grupo OG0000021 formado por proteínas de ligação ao RNA com função de reconhecimento das sequências AU na porção 3'-não traduzida (3'-UTR), porções que controlam vários aspectos da maturação do transcrito, incluindo seleção do sítio de splicing, poliadenilação, estabilidade do RNA, transcrição e abundância dos transcritos (Banerjee et al. 2003; Najafabadi et al. 2013). No gráfico *heatmap* se destacam as ampliações na subfamília leishmaniinae e em *P. confusum*.

O grupo OG0000046 é composto por chaperonas HSP70, proteínas responsáveis pela manutenção da homeostase celular em quase todos os organismos e com uma ampla variedade de funções, muitas vezes altamente adaptadas a determinadas espécies, conforme relatado em protistas parasitas. Nos tripanossomatídeos, as HSP70s desempenham papéis essenciais na adaptação às mudanças ambientais e na diferenciação dos estados induzidos pelo estresse, tornando-as importantes para a transmissão e patogênese (Requena et al. 2015).

O grupo OG0000087 é formado por proteínas motoras chamadas cinesinas que estão envolvidas na remodelação do citoesqueleto e no transporte intracelular. Nos tripanossomatídeos esta proteína sofreu uma grande expansão, resultando em um repertório diversificado de proteínas (An e Li 2018).

Finalmente, o grupo OG0000108 contém proteínas do tipo Ring-box envolvidas na renovação de proteínas em resposta ao dano celular e na progressão do ciclo celular. (Rojas et al. 2017).

Os grupos relacionados à utilização de nutrientes são OG0000010 e OG0000063, que abrigam transportadores do tipo ABC e capturadores de glicerol. Essas proteínas do tipo ABC estão entre as mais antigas e conservadas em todos os reinos. Nos cinetoplastídeos, esta família de transportadores sofreu uma diminuição na sua diversidade filogenética ao evoluir em tripanossomatídeos, porém é comum observar expansões em certas famílias de proteínas, principalmente naquelas famílias que

conferem vantagens adaptativas ao parasita. Verificamos que o grupo de proteínas do tipo ABC é formado pela subfamília do tipo A que está relacionada ao transporte de fosfolipídios e resistência a drogas (Cortés-Selva et al. 2005; da Costa et al. 2018; Leprohon et al. 2006); Além disso, em outros organismos, já foi relatada uma tendência de expansão pela subfamília ABCA. (Gong e Wang 2022; Verrier et al. 2008). Também foi relatado no gráfico *heatmap* o grupo GO0000011, composto por proteínas do tipo ABC envolvidas na resistência a drogas.

O grupo OG0000063 está envolvido na captura do glicerol, composto que, assim como a glicose e a prolina, também é injetado no metabolismo central do carbono (Bringaud et al. 2021). Outras enzimas relacionadas à produção de energia e utilização de nutrientes estão presentes nos grupos OG0000096 e OG0000057. No primeiro grupo está a enzima fumarato redutase que participa da produção de succinato (Mracek et al. 1991), principal produto da glicólise durante o estágio no hospedeiro invertebrado em alguns tripanossomatídeos, assim como o segundo grupo conformado por subunidades ATPase, é fonte de oxidação para a respiração celular.

Em relação aos grupos que contribuem para a interação com os hospedeiros, encontramos diversas funções relacionadas à sinalização celular, resistência a drogas, evasão do sistema imunológico e no processo de infecção. A adenilato ciclase (AC), localizada no grupo OG0000005, participa juntamente com o cAMP no comportamento e diferenciação dos parasitas (Tagoe et al. 2015). Esta proteína está envolvida na sinalização celular e evasão do sistema imunológico, por exemplo, limitando a produção de TNF- α produzido pelas células mielóides como primeira linha de defesa. Os genes AC são organizados em arranjos, onde são expressos diferencialmente dependendo do ciclo celular, condições ambientais e posição dentro desses arranjos (Jackson 2015). Um repertório maior foi encontrado em *T. brucei* em comparação com *T. cruzi* e *Leishmania*, onde são co-expressos junto aos genes VSG, isso devido provavelmente ao fato de ter sofrido uma pressão seletiva positiva, uma vez que este parasita foi exposto por mais tempo ao sistema imunológico e outros ligantes na corrente sanguínea (de Castro Neto et al. 2021).

Os grupos OG0000025 e OG0000076 contém proteínas fosfatases e quinases do tipo serina/treonina dependentes de cálcio. A fosforilação de proteínas pode desempenhar um papel crítico no desenvolvimento dos cinetoplastídeos, sendo observadas nestes organismos mudanças no padrão de fosforilação e expressão de quinases e fosfoquinases durante seus ciclos de vida; por exemplo, em *Trypanosoma* e *Leishmania* proteínas quinases desempenham um papel vital no crescimento e controle da virulência (Orr et al. 2000; Szöör 2010; Vacas et al. 2019).

Outro grupo de quinases é encontrado no grupo OG0000038, formado pelas caseínas quinases 1 (CK1) que são altamente conservadas em eucariotos desde protozoários até humanos. Essas quinases são fundamentais na sinalização celular e muito importantes para a interação

hospedeiro-patógeno, pois participam de diversos processos biológicos, incluindo tráfico de proteínas, regulação do ciclo celular e apoptose, o que resultou em uma localização pleiotropia e interação com diversas estruturas subcelulares. Há muitas evidências de sua associação com doenças infecciosas através da exportação de ortólogos de CK1 para regular as vias de sinalização do hospedeiro, como observado em *Plasmodium* e *Leishmania* (Rachidi et al. 2021). Alguns parálogos dessas proteínas em *Leishmania* têm uma relação regulatória com as chaperonas Hsp70 e Hsp90, sugerindo sua importância na biologia e virulência. Além disso, as variações do número de cópias dos retrotransposons podem ter gerado variações no número de cópias específicas para cada cepa (Rachidi et al. 2021).

Por fim, o gráfico *heatmap* exibe um maior número de cópias nas proteínas com funções que contribuem para a interação do parasita com seu meio externo e com os hospedeiros. O grupo OG0000027 é formado por peptidases calpaína, um tipo de cisteína dependente de cálcio que em tripanossomatídeos experimentaram uma expansão por vezes em combinação com outros domínios de diversas funções e origens, o que gerou uma grande diversidade de famílias que continua sendo regularmente atualizada (Ennes-Vidal et al. 2021). Essas peptidases em *T. cruzi* foram associadas à resistência a condições de estresse (Giese et al. 2008).

O grupo OG0000090 são portadores da família MATE que são proteínas de extrusão multi-antimicrobianas que funcionam como antiportadores de prótons ou drogas/sódio, implicadas na resistência a drogas e amplamente conservadas, englobando bactérias, arqueias e eucariotos (Kabra et al. 2019).

Os grupos que contribuem nos processos de infecção são os grupos OG0000115 e OG0000116. O primeiro é composto por proteínas homólogas à família de receptores de progesterona e adiponectina (PAQRs: *the progesterone and adiponectin family of receptors*) que em *T. cruzi* há evidências da sua funcionalidade e que desencadeia a diferenciação celular, contribuindo para o processo de infecção a macrófagos (Coelho et al. 2021). O segundo grupo é formado por proteínas ecotin-like, um inibidor de serina protease, considerado em diversos patógenos como fator de virulência, incluindo nos gêneros *Trypanosoma* e *Leishmania*, embora não totalmente esclarecido. Essa proteína inibe fortemente as proteases do hospedeiro mamífero, como tripsina, quimotripsina, elastase de neutrófilos (NE) e fatores de coagulação do plasma (Garcia et al. 2020).

Também foram encontrados entre os grupos ortólogos com maior expansão às proteínas GP63, abordado anteriormente, no grupo OG0000006, proteína chave na interação com hospedeiros invertebrados e vertebrados.

Por fim temos o grupo OG0000044 composto por um transportador de cálcio da membrana celular. O cálcio é um mensageiro de sinalização essencial em todas as células eucarióticas, incluindo os tripanossomatídeos, cuja concentração intracelular pode ser letal se não for regulada de forma precisa por diversos mecanismos na membrana celular e nas organelas intracelulares. Em parasitas, este elemento promove mudanças metabólicas vitais como resposta adaptativa rápida e de longa duração contra agentes de estresse ambiental. Particularmente em tripanossomatídeos, foi relacionado ao controle da atividade flagelar, aumento da capacidade de infecção, modulação da proliferação, diferenciação e termoestabilidade para tolerar a temperatura dos hospedeiros. Além disso, o cálcio também é chave na liberação de fatores de virulência como os VSGs em *T. brucei* (Benaim et al. 2020; Docampo e Huang 2015).

Com base em tudo o que foi visto acima, pode-se dizer que houve diminuição da redundância funcional e que a minimização fisiológica é idiossincrática, variando de espécie para espécie, porém a simplificação genética pode ser uma resposta típica em decorrência da evolução como parasitas obrigatórios ou mutualistas. Entre as reduções mais significativas estão as bombas de membrana intracelular e os transportadores ABC. Por outro lado, os parasitas mantiveram e evoluíram genes que proporcionaram vantagens adaptativas, como genes que lhes permitiram aproveitar os nutrientes em seu ambiente, por exemplo, multiplicando transportadores de membrana para aproveitar vários metabólitos (Jackson et al. 2016).

V.9 Filogenômica

Devido a os tripanossomatídeos mostrarem taxas de evolução variadas e muitas vezes rápidas (Briones et al. 1999; Haag et al. 1998; Stevens e Rambaut 2001), se implementou um enfoque de inferência filogenética baseado em modelos de coalescência de múltiplas espécies. Este método toma em consideração separação incompleta de linhagens, por isso é estatisticamente mais consistente do que os métodos usuais baseados na concatenação de um grupo de alinhamentos de múltiplos genes (super matriz).

Esta análise filogenômica utilizou as sequências dos 2803 grupos ortólogos de cópia única obtidos com o programa OrthoFinder (seção V.7) para gerar os alinhamentos necessários (usando o programa MAFFT) para gerar o conjunto de árvores filogenéticas de cada grupo ortólogo de cópia única, que foram obtidas com o programa IqTree. Essas árvores filogenéticas foram submetidas a um método de coalescência de árvores com ASTRAL para inferir a árvore da espécie. Como resultado desta abordagem, obteve-se uma árvore filogenômica robusta (figura 9) e que coincide com a classificação filogenética atual dos principais clados filogenéticos dos tripanossomatídeos (Albanaz et al. 2021; d'Avila-Levy et al. 2020; Durante et al. 2020; Grybchuk et al. 2018; Lukeš et al. 2018).

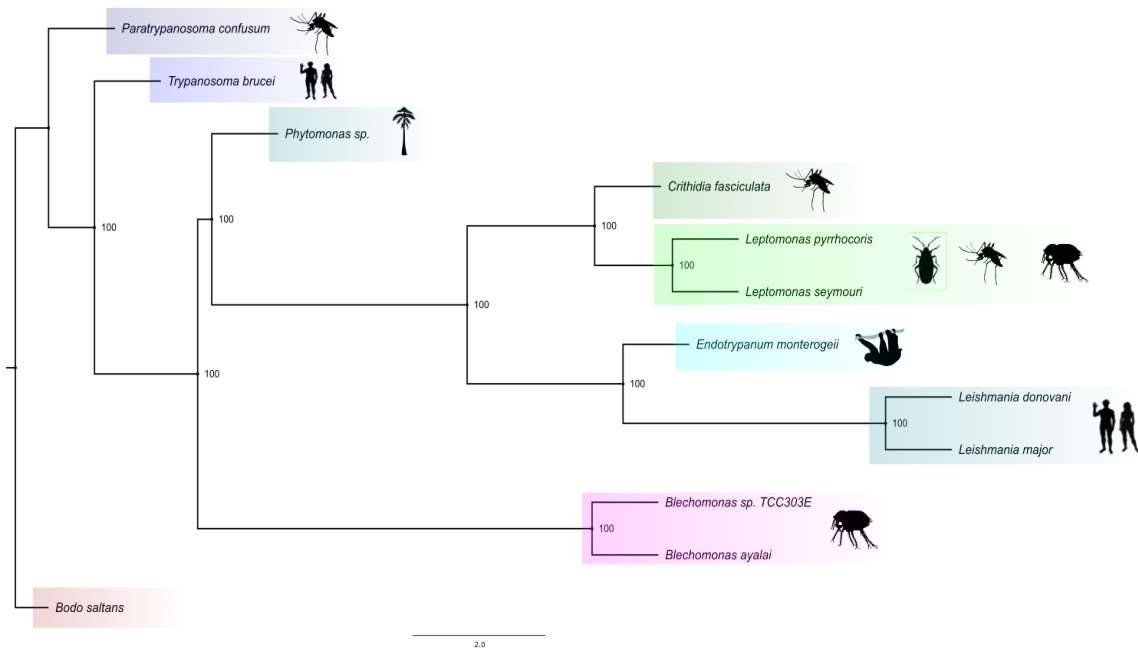


Figura 9. Árvore de espécies (1000 pseudoreplicatas de *bootstrap*) de 2803 grupos ortólogos de cópia única, inferida mediante modelos de coalescência de múltiplas espécies (coalescência de árvores filogenéticas) contendo 12 espécies: 6 espécies monoxênicas, 5 heteroxênicas e como grupo externo, *B. saltans*. *Blechomonas* sp. TCC303E (*Ble*), *B. ayalai* (*Ba*), *C. fasciculata* (*Cf*), *E. monterogeei* (*Em*), *L. donovani* (*Ld*), *L. major* (*Lm*), *L. pyrrocoris* (*Lp*), *L. seymouri* (*Ls*), *T. brucei* (*Tb*), *Phytomonas* AKH_PRJEB1535 (*Ph*), *P. confusum* (*Pc*), *B. saltans* (*Bs*).

O gênero *Blechomonas* juntamente com a subfamília Blechomonadinae foi proposto pela primeira vez em 2013 por (Votýpka et al. 2013) com base na análise filogenética dos genes SL RNA, gGAPDH e SSU rRNA. Este estudo, como vários outros, coloca *Blechomonas* como um clado irmão do restante dos tripanossomatídeos, exceto por Trypanosomatinae (*Trypanosoma* spp.) e Paratrypanosomatinae (*Paratrypanosoma* spp.) que se classificam como os primeiros clados na origem dos tripanossomatídeos. Neste contexto, é inferida uma origem precoce do gênero *Blechomonas* (Grybchuk et al. 2018). De acordo com a literatura, nossos resultados mostram de forma clara e robusta as relações filogenéticas dos principais clados de tripanossomatídeos, com *B. saltans* como grupo externo, seguido na linha evolutiva por *P. confusum* como o tripanossomatídeo mais antigo e mais próximo dos bodonídeos. As espécies de *Blechomonas* estão agrupadas no mesmo clado irmão do restante dos tripanossomatídeos e posterior ao gênero *Trypanosoma*.

V.10. Vias metabólicas

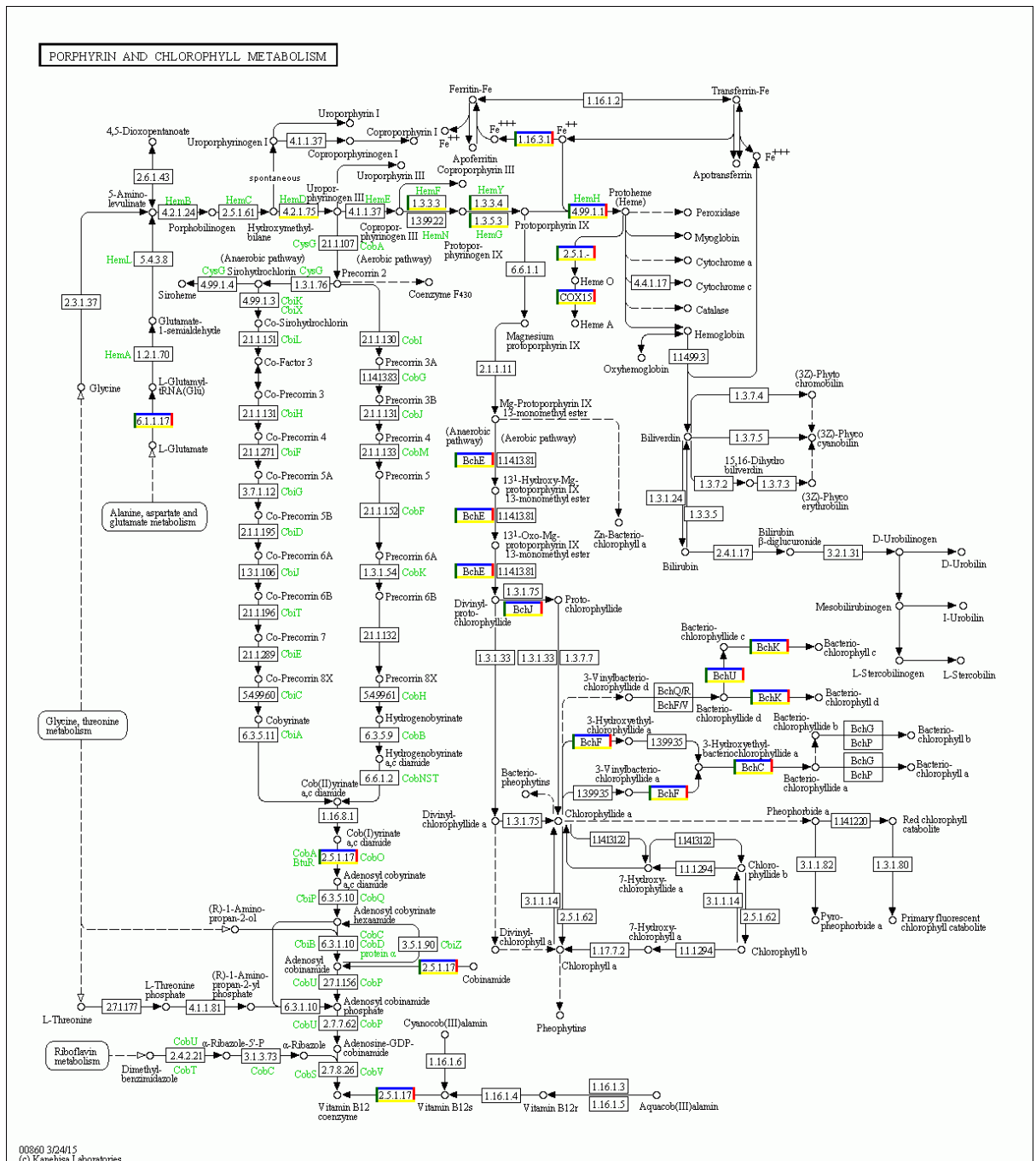
Como visto até agora, os tripanossomatídeos sofreram mudanças importantes em seus repertórios genéticos ao longo de suas histórias evolutivas, como consequência da adaptação aos seus hospedeiros. Notoriamente, eventos de redução genética de certas capacidades metabólicas que não são essenciais para a sobrevivência dentro de seus hospedeiros têm sido observados nesses parasitas,

uma vez que os produtos metabólicos podem ser obtidos dos hospedeiros, como a capacidade de sintetizar grupos heme, purinas e aminoácidos aromáticos, mas também o catabolismo de metabólitos e a degradação de macromoléculas (Jackson 2015; Jackson et al. 2016). Observou-se que *Blechnomonas* segue essa tendência evolutiva de redução e coincide em muitas ocasiões com o resto dos tripanossomatídeos.

V.10.1 Via biossintética do grupo heme

Blechnomonas, como parasita, teve que obter a capacidade de aproveitar o ferro de seu hospedeiro, um metal essencial para a sobrevivência. Neste caso, o genoma de *Blechnomonas* apresentou um gene que codifica uma enzima que reduz o ferro férrico a ferroso e um gene que codifica um transportador de ferro ferroso. No entanto, *Blechnomonas*, como todos os tripanossomatídeos, perdeu a capacidade de sintetizar grupos heme (figura 10), um cofator essencial para muitas enzimas, como os citocromos c. Neste gênero, foi encontrado apenas um gene que codifica uma enzima ferroquelatase (HemH), que catalisa a inserção do ferro ferroso na protoporfirina IX, enquanto em alguns tripanossomatídeos foi observado que eles adquiriram os três últimos genes da via metabólica de formação de grupos heme (Opperdoes e Coombs 2007). Por esta razão, *Blechnomonas* tem que capturar os grupos heme de seus hospedeiros, uma característica observada na maioria dos tripanossomatídeos (Chang et al. 1975; Korený et al. 2010), com exceção de uma minoria que o obtém de seus endossimbiontes bacterianos (Alves et al. 2011). No entanto, neste estudo as anotações genéticas obtidas para os genomas de *Blechnomonas* não indicaram genes codificadores de receptores heme.

Figura 10. Mapa metabólico que mostra a presença e ausência de genes da rota metabólica de biosíntese de moléculas de porfirina de quatro espécies de tripanossomatídeos: *Blechnomonas* TCC303E (linha azul), *B. ayalai* (linha vermelha), *L. major* (linha amarela), *L. pyrrocoris* (linha verde). Mapa obtido da base de dados KEGG e os genes foram atribuídos pelo programa ASGARD.

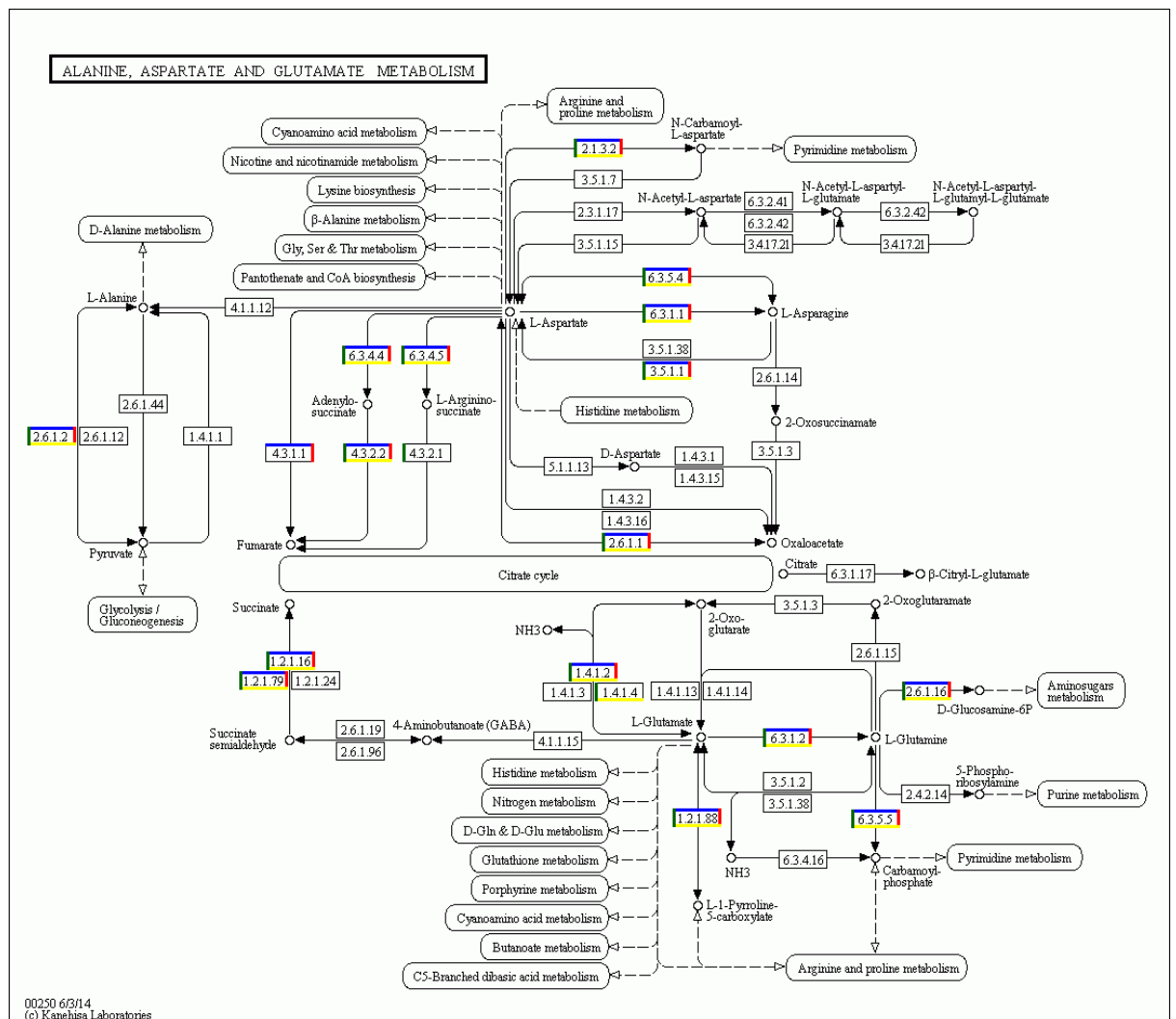


V.10.2 Aminoácidos

A análise da rota metabólica dos aminoácidos mostrada nos mapas metabólicos (figuras 11-14) indicou que o gênero *Blechnomonas*, como o restante dos tripanossomatídeos, é capaz de sintetizar seus próprios aminoácidos não essenciais, incluindo os aminoácidos treonina e metionina

(Opperdoes et al. 2016). Neste gênero observou-se que a alanina pode ser convertida em piruvato pela enzima alanina transaminase (EC 2.6.1.2); este aminoácido é conhecido por estar presente em peptidoglicanos da parede celular de algumas bactérias e provavelmente no intestino médio de insetos (Opperdoes et al. 2016); O fumarato pode ser formado a partir do ácido aspártico, que pode vir da asparagina através da enzima aspartato-amônia ligase (EC 6.3.1.1) ou também pode vir do oxaloacetato através da enzima aspartato aminotransferase (EC 2.6.1.1).

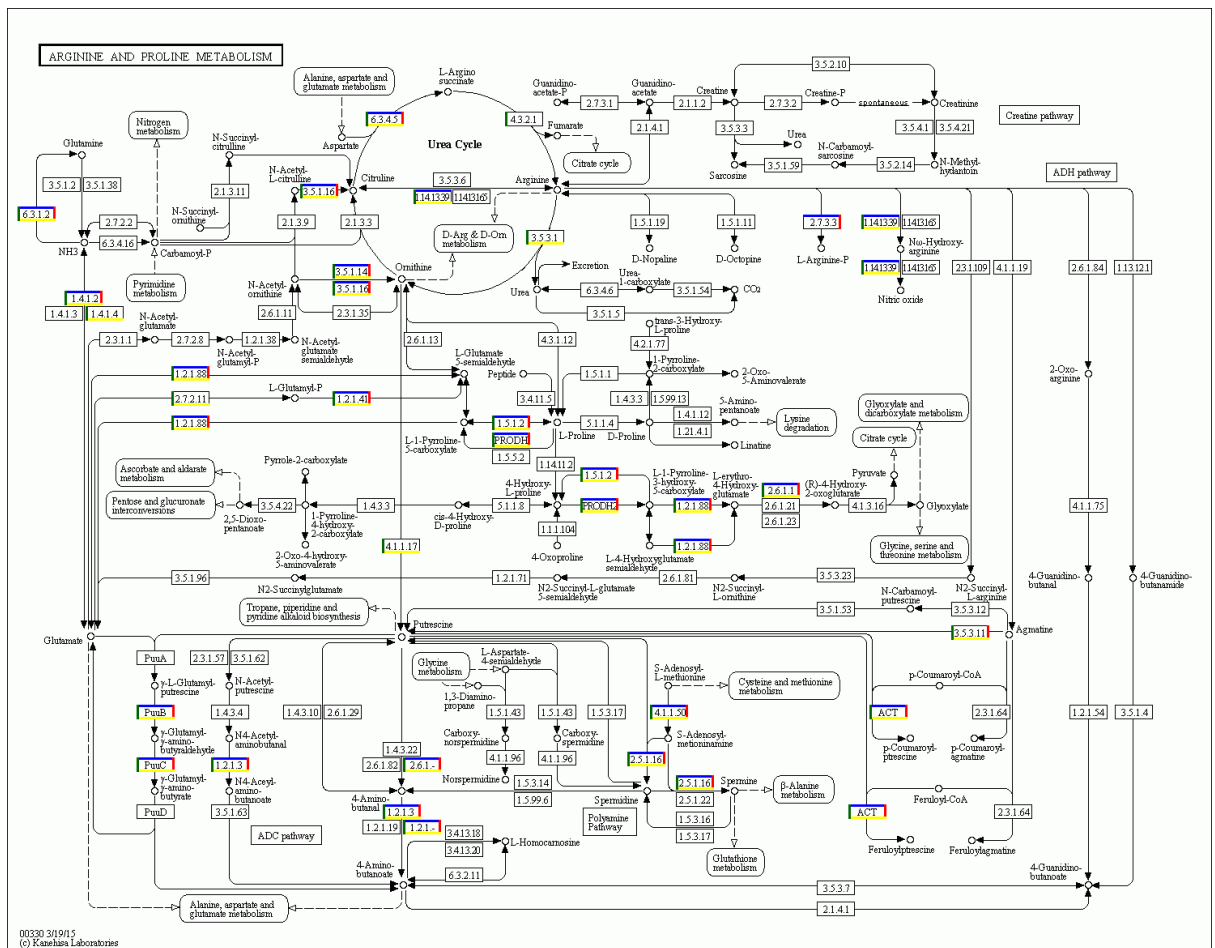
Figura 11. Mapa metabólico que mostra a presença e ausência de genes da rota metabólica dos aminoácidos alanina, ácido aspártico e ácido glutâmico de quatro espécies de tripanossomatídeos: *Blechnomonas* TCC303E (linha azul), *B. ayalai* (linha vermelha), *L. major* (linha amarela), *L. pyrrocoris* (linha verde). Mapa obtido da base de dados KEGG e os genes foram atribuídos pelo programa ASGARD.



Blechnomonas, juntamente com *C. fasciculata* e *Leptomonas* spp, são os únicos capazes de sintetizar arginina (um importante intermediário no ciclo da ureia e predecessor das poliaminas) a partir da citrulina por meio da enzima óxido nítrico sintase (EC 1.14.13.39). A prolina pode ser convertida em ácido glutâmico pelas enzimas pirrolina-5-carboxilato redutase (1.5.1.2) e L-glutamato gama-semialdeído desidrogenase (EC 1.2.1.88), podendo em seguida chegar à produção de glutamina através da glutamato--amônia ligase (EC 6.3.1.2). No entanto, a enzima que produz a reação inversa

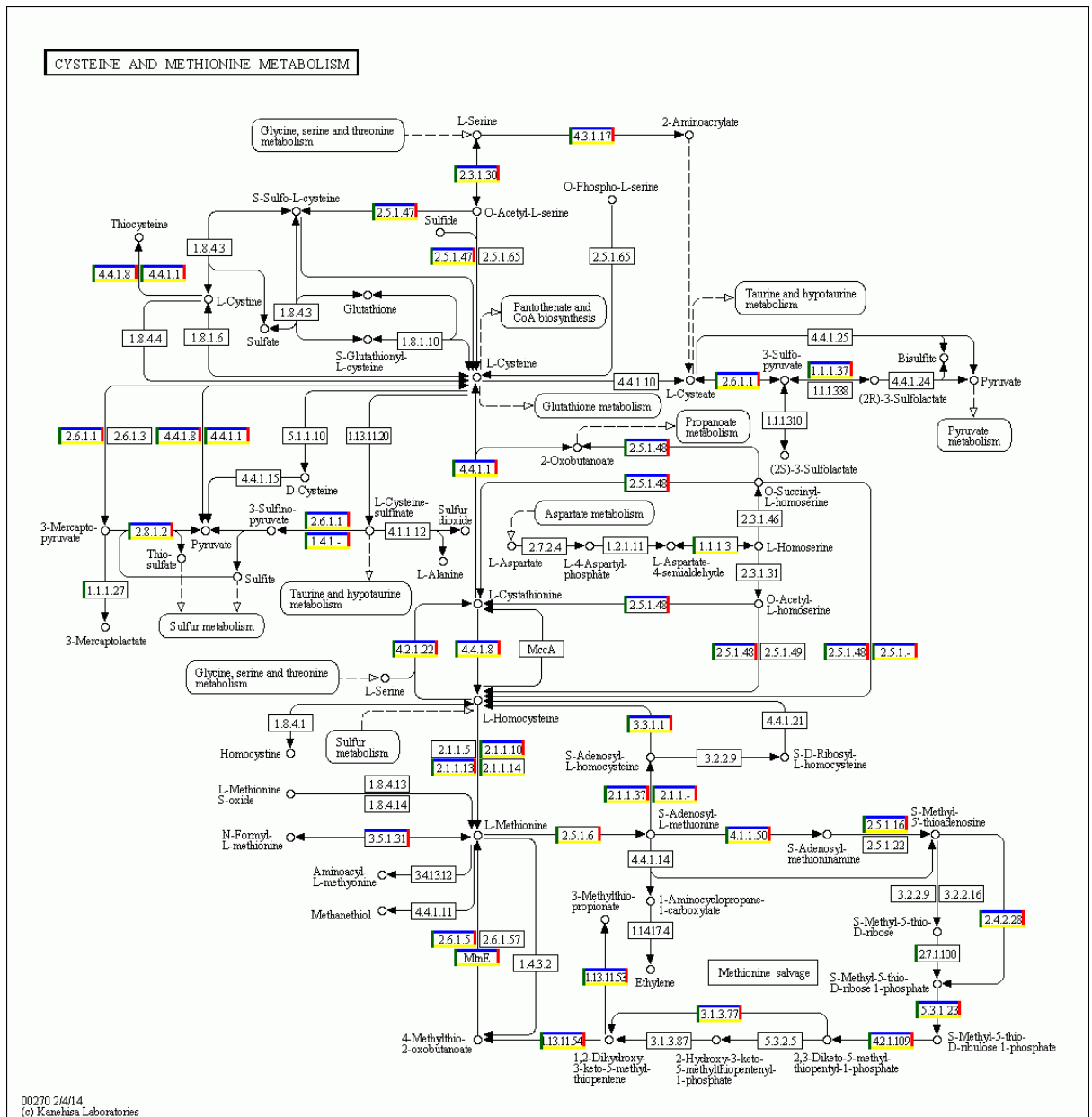
(glutaminase) está ausente, como no restante dos tripanossomatídeos. A reação reversa do ácido glutâmico para a formação de prolina foi observada apenas em organismos do grupo Leishmaniinae, e apenas *B. saltans* é capaz de produzir prolina a partir da ornitina. Estes aminoácidos (prolina e ácido glutâmico) são fonte principal de carbono e amônio para os tripanossomatídeos durante sua estadia no intestino médio dos insetos.

Figura 12. Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos arginina e prolina de quatro espécies de tripanossomatídeos: *Blechnomonas* TCC303E (linha azul), *B. ayalai* (linha vermelha), *L. major* (linha amarela), *L. pyrrocoris* (linha verde). O mapa obtido do banco de dados KEGG e os genes foram atribuídos pelo programa ASgard.



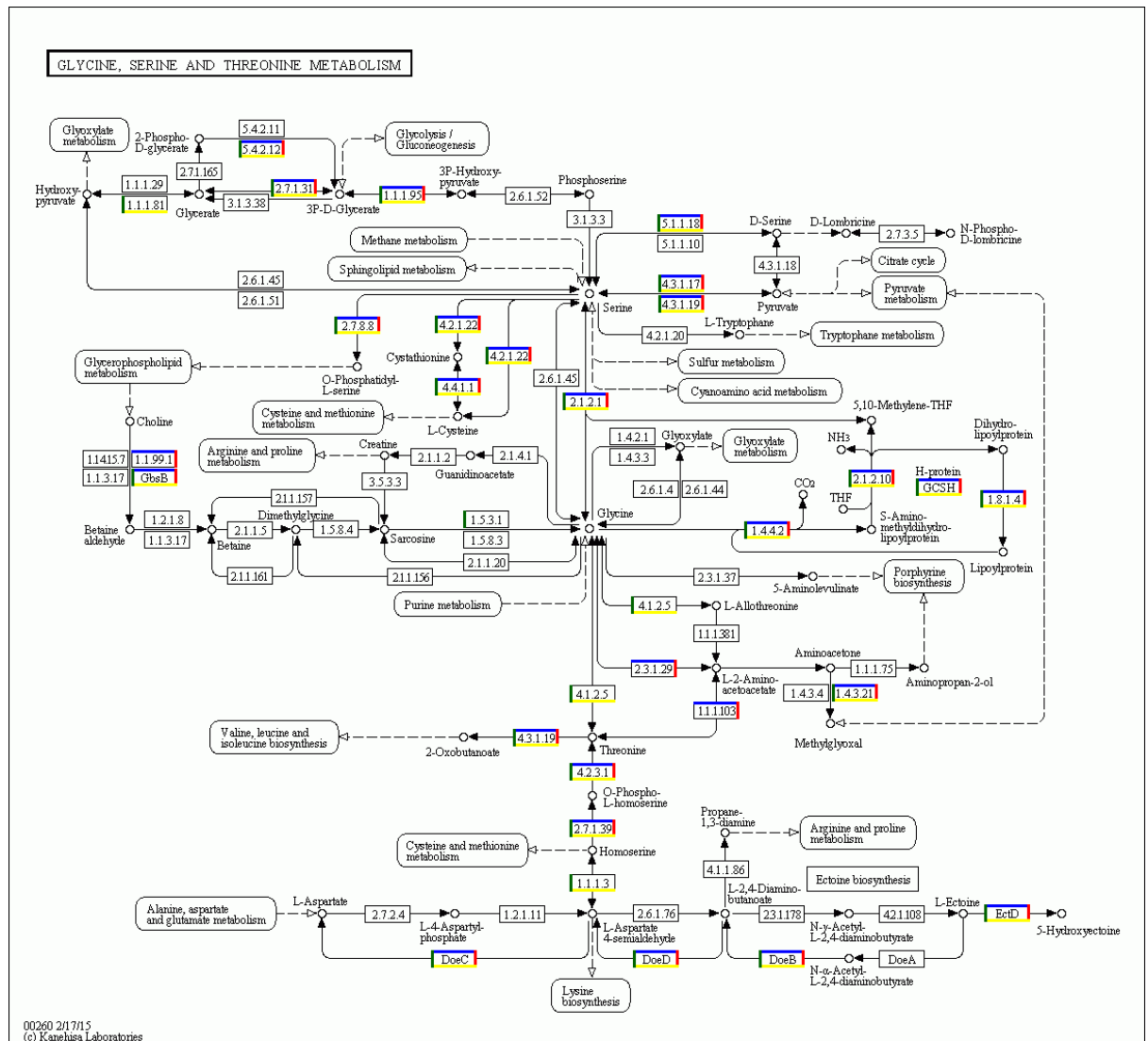
A cisteína pode ser formada a partir da homocisteína pelas enzimas cistationina beta liase (EC 4.4.1.8) e cistationina gama liase (EC 4.4.1.8) ou a partir da serina utilizando a enzima serina O-acetiltransferase (EC 2.3.1.30) e cisteína sintase A (EC 2.5.1.47). A cisteína pode então ser metabolizada em acetil-CoA via mercaptopiruvato. A metionina pode ser formada a partir da serina por várias enzimas (EC 2.3.1.30, EC 2.5.1.47, EC 4.4.1.1, EC 4.4.1.8, EC 2.1.1.10 e EC 2.1.1.13).

Figura 13. Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos cisteína e metionina de quatro espécies de tripanossomatídeos: *Blechnomonas* TCC303E (linha azul), *B. ayalai* (linha vermelha), *L. major* (linha amarela), *L. pyrrohocoris* (linha verde). O mapa obtido do banco de dados KEGG e os genes foram atribuídos pelo programa ASGARD.



A serina pode gerar glicina através da enzima glicina hidroximetiltransferase (EC 2.1.2.1), mas, ao contrário de *L. pyrrohocoris* e *T. brucei*, espécies do gênero *Blechnomonas* não podem formar treonina a partir de glicina.

Figura 14. Mapa metabólico exibindo a presença e ausência dos genes das vias metabólicas dos aminoácidos glicina, serina e metionina de quatro espécies de tripanossomatídeos: *Blechnomonas* TCC303E (linha azul), *B. ayalai* (linha vermelha), *L. major* (linha amarela), *L. pyrrocoris* (linha verde). O mapa obtido do banco de dados KEGG e os genes foram atribuídos pelo programa ASGARD.



Por outro lado, *Blechnomonas*, juntamente com o resto dos tripanossomatídeos, perdeu certas capacidades como catabolizar lisina, triptofano e histidina, e também degradar aerobicamente os aminoácidos aromáticos fenilalanina e tirosina. Além disso, *Blechnomonas*, como *Trypanosoma* e *Phytomonas*, não são capazes de metabolizar os aminoácidos de cadeia ramificada isoleucina e valina, incluindo metionina e treonina, e produzir succinato (intermediário para o ciclo TCA), devido à perda independente de 3 genes da via metil malonil-CoA (propionil-CoA carboxilase, metil-malonil-CoA mutase e metil-malonil-CoA epimerase). Também, eles não podem usar a xilulose pentose porque perderam a xiluloquinase (Oppendoes et al. 2016). Além disso, como em Leishmaniinae, *Blechnomonas* mantém os genes hidroxil-metil-glutaril-CoA (HMG-CoA) liase e β -hidroxi-butirato desidrogenase, enzimas essenciais para a produção de corpos cetônicos a partir da leucina.

V.12 Conclusão

As análises mostraram que *Blechnomonas* sp. TCC303E apresenta um repertório genético e arquitetura genética semelhantes aos dos demais tripanossomatídeos, compartilhando com eles um alto número de grupos ortólogos em seu *core genome*. Isto é refletido numa rede metabólica semelhante, mas também com características genéticas próprias obtidas da sua história evolutiva, especialmente aquelas das famílias multigênicas de superfície, peptidases e transposons que seriam sinal das adaptações ao seu hospedeiro. Finalmente, a filogenômica reafirma uma origem precoce do gênero *Blechnomonas*, como um clado irmão dos tripanossomatídeos, exceto por Trypanosomatinae (*Trypanosoma* spp.) e Paratrypanosomatinae (*Paratrypanosoma* spp.).

REFERÊNCIAS

- Aguiar BG, Dumas C, Maaroufi H, Padmanabhan PK, Papadopoulou B. 2020. The AAA + ATPase valosin-containing protein (VCP)/p97/Cdc48 interaction network in *Leishmania*. *Sci. Rep.* 10(1):13135
- Ahmad M, Tuteja R. 2012. *Plasmodium falciparum* RuvB proteins: Emerging importance and expectations beyond cell cycle progression. *Commun. Integr. Biol.* 5(4):350–61
- Albanaz ATS, Gerasimov ES, Shaw JJ, Sádlová J, Lukeš J, et al. 2021. Genome Analysis of *Endotrypanum* and *Porcisia* spp., Closest Phylogenetic Relatives of *Leishmania*, Highlights the Role of Amastins in Shaping Pathogenicity. *Genes (Basel)*. 12(3):
- Alcolea PJ, Alonso A, García-Tabares F, Toraño A, Larraga V. 2014. An Insight into the proteome of *Crithidia fasciculata* choanomastigotes as a comparative approach to axenic growth, peanut lectin agglutination and differentiation of *Leishmania* spp. promastigotes. *PLoS ONE*. 9(12):e113837
- Alves JMP, Buck GA. 2007. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. *Chem. Biodivers.* 4(11):2593–2602
- Alves JMP, Voegtly L, Matveyev AV, Lara AM, da Silva FM, et al. 2011. Identification and phylogenetic analysis of heme synthesis genes in trypanosomatids and their bacterial endosymbionts. *PLoS ONE*. 6(8):e23518
- An T, Li Z. 2018. An orphan kinesin controls trypanosome morphology transitions by targeting FLAM3 to the flagellum. *PLoS Pathog.* 14(5):e1007101
- Banerjee H, Rahn A, Davis W, Singh R. 2003. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA*. 9(1):88–99
- Barratt J, Kaufer A, Peters B, Craig D, Lawrence A, et al. 2017. Isolation of Novel Trypanosomatid, *Zelonia australiensis* sp. nov. (Kinetoplastida: Trypanosomatidae) Provides Support for a Gondwanan Origin of Dixerous Parasitism in the Leishmaniinae. *PLoS Negl. Trop. Dis.* 11(1):e0005215
- Beavo JA. 1995. Cyclic nucleotide phosphodiesterases: functional implications of multiple isoforms.

Physiol. Rev. 75(4):725–48

- Benaim G, Paniz-Mondolfi AE, Sordillo EM, Martinez-Sotillo N. 2020. Disruption of Intracellular Calcium Homeostasis as a Therapeutic Target Against *Trypanosoma cruzi*. *Front. Cell. Infect. Microbiol.* 10:46
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, et al. 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science.* 309(5733):416–22
- Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29(12):2607–18
- Bringaud F, Ghedin E, El-Sayed NMA, Papadopoulou B. 2008. Role of transposable elements in trypanosomatids. *Microbes Infect.* 10(6):575–81
- Bringaud F, Plazolles N, Pineda E, Asencio C, Villafranz O, et al. 2021. Glycerol, a possible new player in the biology of trypanosomes. *PLoS Pathog.* 17(12):e1010035
- Briones MR, Souto RP, Stolf BS, Zingales B. 1999. The evolution of two *Trypanosoma cruzi* subgroups inferred from rRNA genes can be correlated with the interchange of American mammalian faunas in the Cenozoic and has implications to pathogenicity and host specificity. *Mol. Biochem. Parasitol.* 104(2):219–32
- Brown MJF, Schmid-Hempel R, Schmid-Hempel P. 2003. Strong context-dependent virulence in a host-parasite system: reconciling genetic evidence with theory. *J. Anim. Ecology.* 72(6):994–1002
- Callejas-Hernández F, Rastrojo A, Poveda C, Gironès N, Fresno M. 2018. Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. *Sci. Rep.* 8(1):14631
- Campbell K, Popov V, Soong L. 2004. Identification and molecular characterization of a gene encoding a protective *Leishmania amazonensis* Trp-Asp (WD) protein. *Infect. Immun.* 72(4):2194–2202
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics.* 25(15):1972–73

- Casanova M, Crobu L, Blaineau C, Bourgeois N, Bastien P, Pagès M. 2009. Microtubule-severing proteins are involved in flagellar length control and mitosis in Trypanosomatids. *Mol. Microbiol.* 71(6):1353–70
- Castro Neto AL, Brito ANALM, Rezende AM, Magalhães FB, de Melo Neto OP. 2019. In silico characterization of multiple genes encoding the GP63 virulence protein from *Leishmania braziliensis*: identification of sources of variation and putative roles in immune evasion. *BMC Genomics.* 20(1):118
- Chahar P, Kaushik M, Gill SS, Gakhar SK, Gopalan N, et al. 2015. Genome-Wide Collation of the *Plasmodium falciparum* WDR Protein Superfamily Reveals Malarial Parasite-Specific Features. *PLoS ONE.* 10(6):e0128507
- Chang KP, Chang CS, Sassa S. 1975. Heme biosynthesis in bacterium-protozoon symbioses: enzymic defects in host hemoflagellates and complementary role of their intracellular symbiotes. *Proc Natl Acad Sci USA.* 72(8):2979–83
- Choudhury K, Cardenas D, Pullikuth AK, Catling AD, Aiyar A, Kelly BL. 2011. Trypanosomatid RACK1 orthologs show functional differences associated with translation despite similar roles in *Leishmania* pathogenesis. *PLoS ONE.* 6(6):e20710
- Clayton C. 2019. Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.* 9(6):190072
- Coelho FS, Oliveira MM, Vieira DP, Torres PHM, Moreira ICF, et al. 2021. A novel receptor for platelet-activating factor and lysophosphatidylcholine in *Trypanosoma cruzi*. *Mol. Microbiol.* 116(3):890–908
- Colla SR, Otterstatter MC, Gegear RJ, Thomson JD. 2006. Plight of the bumble bee: Pathogen spillover from commercial to wild populations. *Biol. Conserv.* 129(4):461–67
- Cordeiro-Da-Silva A, Borges MC, Guilvard E, Ouassia A. 2001. Dual role of the *Leishmania major* ribosomal protein S3a homologue in regulation of T- and B-cell activation. *Infect. Immun.* 69(11):6588–96
- Cortés-Selva F, Muñoz-Martínez F, Iliás A, Jiménez AI, Váradi A, et al. 2005. Functional expression of a multidrug P-glycoprotein transporter of *Leishmania*. *Biochem. Biophys. Res. Commun.*

329(2):502–7

- Das S, Shah P, Baharia RK, Tandon R, Khare P, et al. 2013. Over-expression of 60s ribosomal L23a is associated with cellular proliferation in SAG resistant clinical isolates of *Leishmania donovani*. *PLoS Negl. Trop. Dis.* 7(12):e2527
- d’Avila-Levy CM, Altoé ECF, Uehara LA, Branquinha MH, Santos ALS. 2014. GP63 function in the interaction of trypanosomatids with the invertebrate host: facts and prospects. *Subcell Biochem.* 74:253–70
- d’Avila-Levy CM, Bearzatto B, Ambroise J, Helaers R, Butenko A, et al. 2020. First Draft Genome of the Trypanosomatid *Herpetomonas muscarum ingenoplastis* through MinION Oxford Nanopore Technology and Illumina Sequencing. *TropicalMed.* 5(1):
- da Costa KM, Valente RC, Salustiano EJ, Gentile LB, Freire-de-Lima L, et al. 2018. Functional Characterization of ABCC Proteins from *Trypanosoma cruzi* and Their Involvement with Thiol Transport. *Front. Microbiol.* 9:205
- Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics.* 23(6):673–79
- de Castro Neto AL, da Silveira JF, Mortara RA. 2021. Comparative analysis of virulence mechanisms of trypanosomatids pathogenic to humans. *Front. Cell. Infect. Microbiol.* 11:669079
- Diniz MC, Pacheco ACL, Girão KT, Araujo FF, Walter CA, Oliveira DM. 2010. The tetratricopeptide repeats (TPR)-like superfamily of proteins in *Leishmania* spp., as revealed by multi-relational data mining. *Pattern Recognit. Lett.* 31(14):2178–89
- Docampo R, Huang G. 2015. Calcium signaling in trypanosomatid parasites. *Cell Calcium.* 57(3):194–202
- Durante IM, Butenko A, Rašková V, Charyyeva A, Svobodová M, et al. 2020. Large-Scale Phylogenetic Analysis of Trypanosomatid Adenylate Cyclases Reveals Associations with Extracellular Lifestyle and Host-Pathogen Interplay. *Genome Biol. Evol.* 12(12):2403–16
- El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, et al. 2005a. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science.* 309(5733):409–15
- El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, et al. 2005b. Comparative genomics of

- trypanosomatid parasitic protozoa. *Science*. 309(5733):404–9
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20(1):238
- Ennes-Vidal V, Branquinha MH, Dos Santos ALS, d'Avila-Levy CM. 2021. The diverse calpain family in trypanosomatidae: functional proteins devoid of proteolytic activity? *Cells*. 10(2):
- Field MC, Natesan SKA, Gabernet-Castello C, Koumandou VL. 2007. Intracellular trafficking in the trypanosomatids. *Traffic*. 8(6):629–39
- Field MC. 2005. Signalling the genome: the Ras-like small GTPase family of trypanosomatids. *Trends Parasitol.* 21(10):447–50
- Flegontov P, Butenko A, Firsov S, Kraeva N, Eliáš M, et al. 2016. Genome of *Leptomonas pyrrocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci. Rep.* 6:23704
- Garcia FB, Cabral AD, Fuhlendorf MM, da Cruz GF, Dos Santos JV, et al. 2020. Functional and structural characterization of an ecotin-like serine protease inhibitor from *Trypanosoma cruzi*. *Int. J. Biol. Macromol.* 151:459–66
- Ghedini E, Bringaud F, Peterson J, Myler P, Berriman M, et al. 2004. Gene synteny and evolution of genome architecture in trypanosomatids. *Mol. Biochem. Parasitol.* 134(2):183–91
- Giese V, Dallagiovanna B, Marchini FK, Pavoni DP, Krieger MA, Goldenberg S. 2008. *Trypanosoma cruzi*: a stage-specific calpain-like protein is induced after various kinds of stress. *Mem. Inst. Oswaldo Cruz*. 103(6):598–601
- Glisovic T, Bachorik JL, Yong J, Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 582(14):1977–86
- Gong X, Wang S. 2022. New Insights into Evolution of the ABC Transporter Family in *Mesostigma viride*, a Unicellular Charophyte Algae. *Curr. Issues Mol. Biol.* 44(4):1646–60
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res.* 31(1):439–41
- Grisard EC, Stoco PH, Wagner G, Sincero TCM, Rotava G, et al. 2010. Transcriptomic analyses of the avirulent protozoan parasite *Trypanosoma rangeli*. *Mol. Biochem. Parasitol.* 174(1):18–25

- Grybchuk D, Kostygov AY, Macedo DH, Votýpka J, Lukeš J, Yurchenko V. 2018. RNA viruses in blechomonas (trypanosomatidae) and evolution of leishmanivirus. *MBio*. 9(5):
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29(8):1072–75
- Haag J, O’huigin C, Overath P. 1998. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Mol. Biochem. Parasitol*. 91(1):37–49
- Hammarton TC, Mottram JC, Doerig C. 2003. The cell cycle of parasitic protozoa: potential for chemotherapeutic exploitation. *Prog. Cell Cycle Res*. 5:91–101
- Hendrickx S, Reis-Cunha JL, Forrester S, Jeffares DC, Caljon G. 2021. Experimental Selection of Paromomycin Resistance in *Leishmania donovani* Amastigotes Induces Variable Genomic Polymorphisms. *Microorganisms*. 9(8):
- Hoare CA, Wallace FG. 1966. Developmental stages of trypanosomatid flagellates: a new terminology. *Nature*. 212(5068):1385–86
- Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, et al. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res*. 44(D1):D286-93
- Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, et al. 2005. The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 309(5733):436–42
- Jackson AP, Otto TD, Aslett M, Armstrong SD, Bringaud F, et al. 2016. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr. Biol*. 26(2):161–72
- Jackson AP. 2015. Genome evolution in trypanosomatid parasites. *Parasitology*. 142 Suppl 1:S40-56
- Jensen RE, Englund PT. 2012. Network news: the replication of kinetoplast DNA. *Annu. Rev. Microbiol*. 66:473–91
- Jiang DW, Englund PT. 2001. Four *Trypanosoma brucei* fatty acyl-CoA synthetases: fatty acid specificity of the recombinant proteins. *Biochem. J*. 358(Pt 3):757–61
- Jirků M, Yurchenko VY, Lukeš J, Maslov DA. 2012. New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. *J. Eukaryot*.

Microbiol. 59(6):537–47

- Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics.* 11:431
- Kabra R, Chauhan N, Kumar A, Ingale P, Singh S. 2019. Efflux pumps and antimicrobial resistance: Paradoxical components in systems genomics. *Prog. Biophys. Mol. Biol.* 141:15–24
- Kaiser A. 2019. Druggable Targets in Cyclic Nucleotide Signaling Pathways in Apicomplexan Parasites and Kinetoplastids against Disabling Protozoan Diseases in Humans. *Int. J. Mol. Sci.* 20(1):
- Kalesh K, Sundriyal S, Perera H, Cobb SL, Denny PW. 2021. Quantitative Proteomics Reveals that Hsp90 Inhibition Dynamically Regulates Global Protein Synthesis in *Leishmania mexicana*. *mSystems.* 6(3):
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30(4):772–80
- Kaufer A, Ellis J, Stark D, Barratt J. 2017. The evolution of trypanosomatid taxonomy. *Parasit. Vectors.* 10(1):287
- Kaufer A, Stark D, Ellis J. 2019. Evolutionary Insight into the Trypanosomatidae Using Alignment-Free Phylogenomics of the Kinetoplast. *Pathogens.* 8(3):
- Kedzierski L, Montgomery J, Bullen D, Curtis J, Gardiner E, et al. 2004. A leucine-rich repeat motif of *Leishmania* parasite surface antigen 2 binds to macrophages through the complement receptor 3. *J. Immunol.* 172(8):4902–6
- Koonin EV, Galperin MY. 2003. Principles and methods of sequence analysis. In *Sequence — Evolution — Function*, pp. 111–92. Boston, MA: Springer US
- Korený L, Lukes J, Oborník M. 2010. Evolution of the haem synthetic pathway in kinetoplastid flagellates: an essential pathway that is not essential after all? *Int. J. Parasitol.* 40(2):149–56
- Kostygov AY, Yurchenko V. 2017. Revised classification of the subfamily Leishmaniinae (Trypanosomatidae). *Folia Parasitol.* 64:
- Kraeva N, Butenko A, Hlaváčová J, Kostygov A, Myšková J, et al. 2015. *Leptomonas seymouri*: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome

- Profiling and Co-infection with *Leishmania donovani*. *PLoS Pathog.* 11(8):e1005127
- Kulkarni PG, Shah N, Waghela BN, Pathak CM, Pappachan A. 2019. *Leishmania donovani* adenylate kinase 2a prevents ATP-mediated cell cytolysis in macrophages. *Parasitol. Int.* 72:101929
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. 2004. Versatile and open software for comparing large genomes. *Genome Biol.* 5(2):R12
- Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35(9):3100–3108
- Leprohon P, Légaré D, Girard I, Papadopoulou B, Ouellette M. 2006. Modulation of *Leishmania* ABC protein gene expression through life stages and among drug-resistant parasites. *Eukaryotic Cell.* 5(10):1713–25
- Linardi PM. 2017. Fleas and Diseases. In *Arthropod Borne Diseases*, ed. CB Marcondes, pp. 517–36. Cham: Springer International Publishing
- Linhares-Lacerda L, Morrot A. 2016. Role of small rnas in trypanosomatid infections. *Front. Microbiol.* 7:367
- Li Q, Zhao P, Li J, Zhang C, Wang L, Ren Z. 2014. Genome-wide analysis of the WD-repeat protein family in cucumber and *Arabidopsis*. *Mol. Genet. Genomics.* 289(1):103–24
- Llanes A, Restrepo CM, Del Vecchio G, Anguizola FJ, Lleonart R. 2015. The genome of *Leishmania panamensis*: insights into genomics of the *L. (Viannia)* subgenus. *Sci. Rep.* 5:8550
- Llewellyn MS, Messenger LA, Luquetti AO, Garcia L, Torrico F, et al. 2015. Deep sequencing of the *Trypanosoma cruzi* GP63 surface proteases reveals diversity and diversifying selection among chronic and congenital Chagas disease patients. *PLoS Negl. Trop. Dis.* 9(4):e0003458
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–64
- 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):466–80
- Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. 2014. Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.* 195(2):115–22

- Maslov DA, Podlipaev SA, Lukes J. 2001. Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. *Mem. Inst. Oswaldo Cruz.* 96(3):397–402
- Maslov DA, Votýpka J, Yurchenko V, Lukeš J. 2013. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol.* 29(1):43–52
- Maslov DA, Westenberger SJ, Xu X, Campbell DA, Sturm NR. 2007. Discovery and barcoding by analysis of spliced leader RNA gene sequences of new isolates of Trypanosomatidae from Heteroptera in Costa Rica and Ecuador. *J. Eukaryot. Microbiol.* 54(1):57–65
- McGhee RB, Cosgrove WB. 1980. Biology and physiology of the lower Trypanosomatidae. *Microbiol. Rev.* 44(1):140–73
- Medina-Acosta E, Paul S, Tomlinson S, Pontes-de-Carvalho LC. 1994. Combined occurrence of trypanosomal sialidase/trans-sialidase activities and leishmanial metalloproteinase gene homologues in *Endotrypanum* sp. *Mol. Biochem. Parasitol.* 64(2):273–82
- Merritt C, Silva LE, Tanner AL, Stuart K, Pollastri MP. 2014. Kinases as druggable targets in trypanosomatid protozoan parasites. *Chem. Rev.* 114(22):11280–304
- Michels PAM, Villafraz O, Pineda E, Alencar MB, Cáceres AJ, et al. 2021. Carbohydrate metabolism in trypanosomatids: New insights revealing novel complexity, diversity and species-unique features. *Exp. Parasitol.* 224:108102
- Morio F, Reynes J, Dollet M, Pratlong F, Dedet J-P, Ravel C. 2008. Isolation of a protozoan parasite genetically related to the insect trypanosomatid *Herpetomonas samuelpessoai* from a human immunodeficiency virus-positive patient. *J. Clin. Microbiol.* 46(11):3845–47
- Mracek J, Snyder SJ, Chavez UB, Turrens JF. 1991. A soluble fumarate reductase in *Trypanosoma brucei* procyclic trypomastigotes. *J. Protozool.* 38(6):554–58
- Najafabadi HS, Lu Z, MacPherson C, Mehta V, Adoue V, et al. 2013. Global identification of conserved post-transcriptional regulatory programs in trypanosomatids. *Nucleic Acids Res.* 41(18):8591–8600
- Naora H. 1999. Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity? *Immunol. Cell Biol.* 77(3):197–205

- Natesan SKA, Peacock L, Leung KF, Matthews KR, Gibson W, Field MC. 2009. The trypanosome Rab-related proteins RabX1 and RabX2 play no role in intracellular trafficking but may be involved in fly infectivity. *PLoS ONE*. 4(9):e7217
- Naula C, Parsons M, Mottram JC. 2005. Protein kinases as drug targets in trypanosomes and Leishmania. *Biochim. Biophys. Acta*. 1754(1–2):151–59
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics*. 29(22):2933–35
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32(1):268–74
- Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J. 2016. Comparative Metabolism of Free-living Bodo saltans and Parasitic Trypanosomatids. *J. Eukaryot. Microbiol.* 63(5):657–78
- Opperdoes FR, Coombs GH. 2007. Metabolism of Leishmania: proven and predicted. *Trends Parasitol.* 23(4):149–58
- Orr GA, Werner C, Xu J, Bennett M, Weiss LM, et al. 2000. Identification of novel serine/threonine protein phosphatases in Trypanosoma cruzi: a potential role in control of cytokinesis and morphology. *Infect. Immun.* 68(3):1350–58
- Pacheco RS, Marzochi MC, Pires MQ, Brito CM, Madeira M de F, Barbosa-Santos EG. 1998. Parasite genotypically related to a monoxenous trypanosomatid of dog's flea causing opportunistic infection in an HIV positive patient. *Mem. Inst. Oswaldo Cruz*. 93(4):531–37
- Parsons M, Worthey EA, Ward PN, Mottram JC. 2005. Comparative analysis of the kinomes of three pathogenic trypanosomatids: Leishmania major, Trypanosoma brucei and Trypanosoma cruzi. *BMC Genomics*. 6:127
- Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, et al. 2012. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. *PLoS Biol.* 10(11):e1001419
- Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, et al. 2007. Comparative genomic analysis of three Leishmania species that cause diverse human disease. *Nat. Genet.* 39(7):839–47
- Picchi GFA, Zulkiewicz V, Krieger MA, Zanchin NT, Goldenberg S, de Godoy LMF. 2017.

- Post-translational Modifications of Trypanosoma cruzi Canonical and Variant Histones. *J. Proteome Res.* 16(3):1167–79
- Pita S, Díaz-Viraqué F, Iraola G, Robello C. 2019. The tritryps comparative repeatome: insights on repetitive element evolution in trypanosomatid pathogens. *Genome Biol. Evol.* 11(2):546–51
- Plischuk S, Lange CE. 2009. Invasive *Bombus terrestris* (Hymenoptera: Apidae) parasitized by a flagellate (Euglenozoa: Kinetoplastea) and a neogregarine (Apicomplexa: Neogregarinorida). *J. Invertebr. Pathol.* 102(3):263–65
- Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33(Database issue):D501-4
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, et al. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40(Database issue):D290-301
- Rachidi N, Knippschild U, Späth GF. 2021. Dangerous duplicity: the dual functions of casein kinase 1 in parasite biology and host subversion. *Front. Cell. Infect. Microbiol.* 11:655700
- Rajan KS, Chikne V, Decker K, Waldman Ben-Asher H, Michaeli S. 2019. Unique Aspects of rRNA Biogenesis in Trypanosomatids. *Trends Parasitol.* 35(10):778–94
- Rangel DA, Lisboa CV, Novaes RLM, Silva BA, Souza R de F, et al. 2019. Isolation and characterization of trypanosomatids, including *Crithidia mellificae*, in bats from the Atlantic Forest of Rio de Janeiro, Brazil. *PLoS Negl. Trop. Dis.* 13(7):e0007527
- Requena JM, Montalvo AM, Fraga J. 2015. Molecular Chaperones of Leishmania: Central Players in Many Stress-Related and -Unrelated Physiological Processes. *Biomed Res. Int.* 2015:301326
- Rojas F, Koszela J, Búa J, Llorente B, Burchmore R, et al. 2017. The ubiquitin-conjugating enzyme CDC34 is essential for cytokinesis in contrast to putative subunits of a SCF complex in *Trypanosoma brucei*. *PLoS Negl. Trop. Dis.* 11(6):e0005626
- Romagnoli BAA, Holetz FB, Alves LR, Goldenberg S. 2020. RNA Binding Proteins and Gene Expression Regulation in *Trypanosoma cruzi*. *Front. Cell. Infect. Microbiol.* 10:56
- Romaniuk MA, Cervini G, Cassola A. 2016. Regulation of RNA binding proteins in trypanosomatid protozoan parasites. *World J. Biol. Chem.* 7(1):146–57

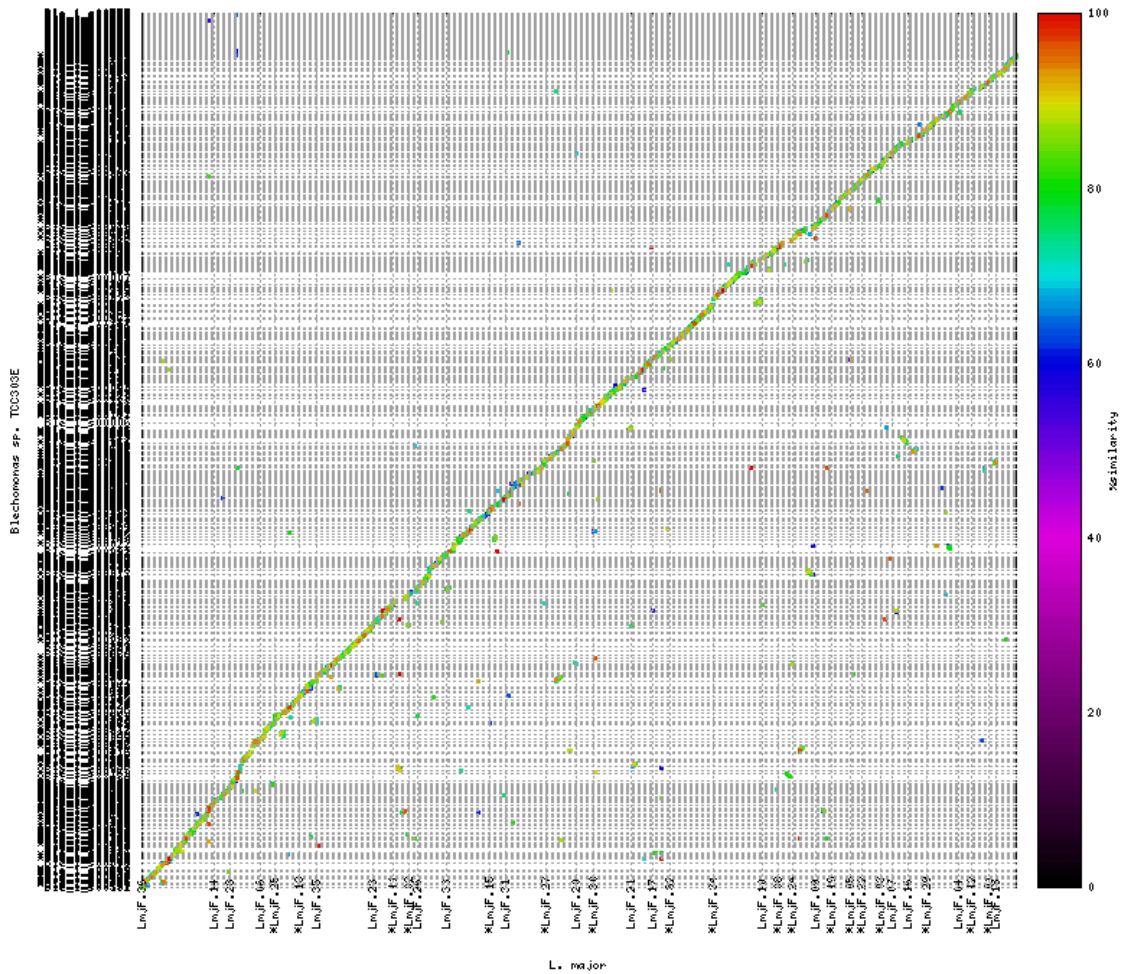
- Santos ALS, Branquinha MH, D'Avila-Levy CM. 2006. The ubiquitous gp63-like metalloprotease from lower trypanosomatids: in the search for a function. *An. Acad. Bras. Ciênc.* 78(4):687–714
- Schaub GA. 1994. Pathogenicity of trypanosomatids on insects. *Parasitology Today.* 10(12):463–68
- Schaub GA. 2009. Chapter 4 interactions of trypanosomatids and triatomines. In Vol. 37, pp. 177–242. Elsevier
- Schmid-Hempel P, Aebi M, Barribeau S, Kitajima T, du Plessis L, et al. 2018. The genomes of *Crithidia bombi* and *C. expoeki*, common parasites of bumblebees. *PLoS ONE.* 13(1):e0189738
- Seppy M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962:227–45
- Silber AM, Pereira CA. 2012. Assignment of putative functions to membrane “hypothetical proteins” from the *Trypanosoma cruzi* genome. *J. Membr. Biol.* 245(3):125–29
- Silvestre A, Shintre SS, Rachidi N. 2022. Released Parasite-Derived Kinases as Novel Targets for Antiparasitic Therapies. *Front. Cell. Infect. Microbiol.* 12:825458
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31(19):3210–12
- Simpson AGB, Stevens JR, Lukes J. 2006. The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol.* 22(4):168–74
- Singh N, Chikara S, Sundar S. 2013. SOLiD™ sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm leptomonas co-infection and raise some key questions. *PLoS ONE.* 8(2):e55738
- Sloan MA, Brooks K, Otto TD, Sanders MJ, Cotton JA, Ligoxygakis P. 2019. Transcriptional and genomic parallels between the monoxenous parasite *Herpetomonas muscarum* and *Leishmania*. *PLoS Genet.* 15(11):e1008452
- Smith TK, Bütikofer P. 2010. Lipid metabolism in *Trypanosoma brucei*. *Mol. Biochem. Parasitol.* 172(2):66–79
- Snider J, Thibault G, Houry WA. 2008. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* 9(4):216

- Souza W de. 2008. Electron microscopy of trypanosomes--a historical view. *Mem. Inst. Oswaldo Cruz.* 103(4):313–25
- Stevens JR, Noyes HA, Dover GA, Gibson WC. 1999. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology.* 118 (Pt 1):107–16
- Stevens J, Rambaut A. 2001. Evolutionary rate differences in trypanosomes. *Infect. Genet. Evol.* 1(2):143–50
- Sylvia O, Aldo S. 2019. Excavata-Kinetoplastea Trypanosomatidae Parasites and the Interaction with their Hosts. *Int. J. Trop. Dis.* 2(1):
- Szöör B. 2010. Trypanosomatid protein phosphatases. *Mol. Biochem. Parasitol.* 173(2):53–63
- Tagoe DNA, Kalejaiye TD, de Koning HP. 2015. The ever unfolding story of cAMP signaling in trypanosomatids: vive la difference! *Front. Pharmacol.* 6:185
- Thomas MC, Macias F, Alonso C, López MC. 2010. The biology and evolution of transposable elements in parasites. *Trends Parasitol.* 26(7):350–62
- Tschoeke DA, Nunes GL, Jardim R, Lima J, Dumaresq AS, et al. 2014. The Comparative Genomics and Phylogenomics of *Leishmania amazonensis* Parasite. *Evol. Bioinform. Online.* 10:131–53
- Ullman B. 1995. Multidrug resistance and P-glycoproteins in parasitic protozoa. *J. Bioenerg. Biomembr.* 27(1):77–84
- Vacas A, Fernández-Rubio C, Algarabel M, Peña-Guerrero J, Larrea E, et al. 2019. The Novel Serine/Threonine Protein Kinase LmjF.22.0810 from *Leishmania major* may be Involved in the Resistance to Drugs such as Paromomycin. *Biomolecules.* 9(11):
- van Nocker S, Ludwig P. 2003. The WD-repeat protein superfamily in Arabidopsis: conservation and divergence in structure and function. *BMC Genomics.* 4(1):50
- Verrier PJ, Bird D, Burla B, Dassa E, Forestier C, et al. 2008. Plant ABC proteins--a unified nomenclature and updated inventory. *Trends Plant Sci.* 13(4):151–59
- Vickerman K. 1994. The evolutionary expansion of the trypanosomatid flagellates. *Int. J. Parasitol.* 24(8):1317–31
- Vij A, Biswas A, Bhattacharya A, Das PK. 2014. A soluble phosphodiesterase in *Leishmania*

- donovani negatively regulates cAMP signaling by inhibiting protein kinase A through a two way process involving catalytic as well as non-catalytic sites. *Int. J. Biochem. Cell Biol.* 57:197–206
- Votýpka J, Klepetková H, Jirků M, Kment P, Lukeš J. 2012a. Phylogenetic relationships of trypanosomatids parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. *Int. J. Parasitol.* 42(5):489–500
- Votýpka J, Klepetková H, Yurchenko VY, Horák A, Lukeš J, Maslov DA. 2012b. Cosmopolitan distribution of a trypanosomatid *Leptomonas pyrrocoris*. *Protist.* 163(4):616–31
- Votýpka J, Suková E, Kraeva N, Ishemgulova A, Duží I, et al. 2013. Diversity of trypanosomatids (Kinetoplastea: Trypanosomatidae) parasitizing fleas (Insecta: Siphonaptera) and description of a new genus *Blechomonas* gen. n. *Protist.* 164(6):763–81
- Wallace FG. 1966. The trypanosomatid parasites of insects and arachnids. *Exp. Parasitol.* 18(1):124–93
- Wendler P, Ciniawsky S, Kock M, Kube S. 2012. Structure and function of the AAA+ nucleotide binding pocket. *Biochim. Biophys. Acta.* 1823(1):2–14
- Wheeler RJ, Gluenz E, Gull K. 2013. The limits on trypanosomatid morphological diversity. *PLoS ONE.* 8(11):e79581
- Wiese M, Kuhn D, Grünfelder CG. 2003. Protein kinase involved in flagellar-length control. *Eukaryotic Cell.* 2(4):769–77
- Yao C. 2010. Major surface protease of trypanosomatids: one size fits all? *Infect. Immun.* 78(1):22–31
- Yurchenko V, Kostygov A, Havlová J, Grybchuk-Ieremenko A, Ševčíková T, et al. 2016. Diversity of Trypanosomatids in Cockroaches and the Description of *Herpetomonas tarakana* sp. n. *J. Eukaryot. Microbiol.* 63(2):198–209
- Zhang C, Sayyari E, Mirarab S. 2017. ASTRAL-III: Increased Scalability and Impacts of Contracting Low Support Branches. In *Comparative Genomics*, Vol. 10562, eds. J Meidanis, L Nakhleh, pp. 53–75. Cham: Springer International Publishing
- Zhang N, Jiang N, Zhang K, Zheng L, Zhang D, et al. 2020. Landscapes of protein posttranslational modifications of african trypanosoma parasites. *iScience.* 23(5):101074
- Zheng Y, Cai X, Bradley JE. 2013. microRNAs in parasites and parasite infection. *RNA Biol.*

10(3):371-79

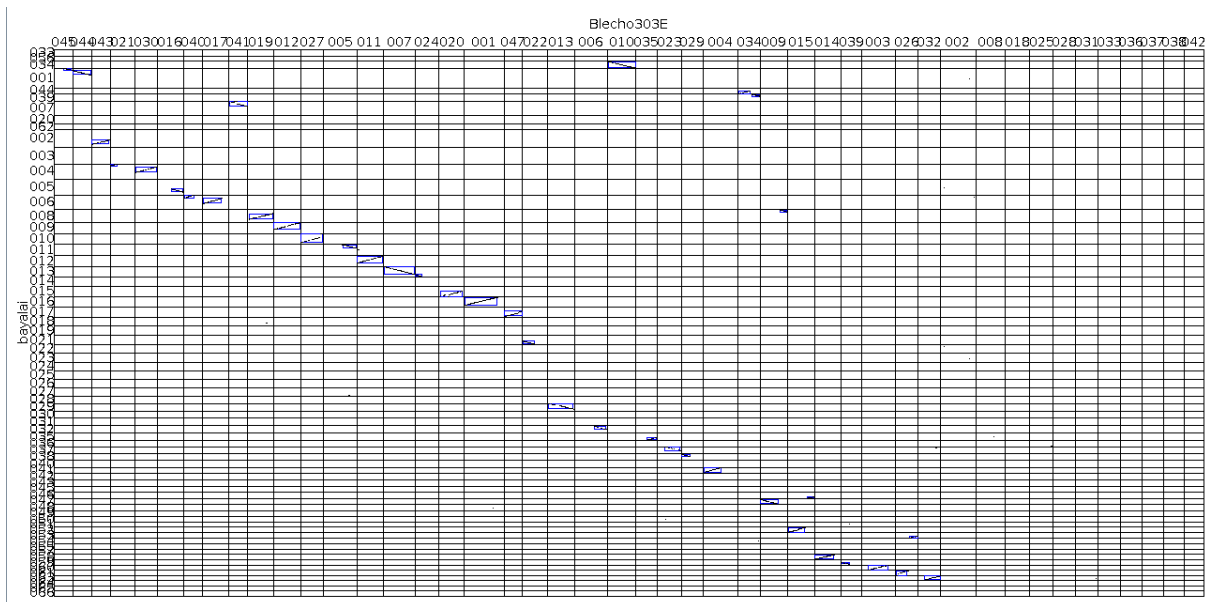
ANEXO



Anexo A: Gráfico *dotplot* gerado pelo pacote MUMmerplot do resultado do programa PROmer para a comparação estrutural no nível de sequências de aminoácidos da montagem de *Blechomonas sp. TCC303E* com o genoma de *L. major*. No eixo horizontal estão distribuídos todos os cromossomos (36) do genoma de referência, *L. major*, e no eixo vertical estão todos os *scaffolds* (728) da montagem do genoma de *Blechomonas sp. TCC303E*. As cores representam porcentagens de similaridade nos alinhamentos. A diagonal mostra uma alta sintonia entre os genomas comparados.

Anexo B. Quantidades dos RNAs de transferência para a ligação a todos os tipos de aminoácidos (71 tRNAs) no genoma de *Blechnomonas* sp. TCC303E.

| Aminoácidos | Números de tRNAs |
|-----------------|------------------|
| Alanina | 7 |
| Arginina | 7 |
| Asparagina | 3 |
| Ácido aspártico | 4 |
| Cisteína | 3 |
| Fenilalanina | 1 |
| Glicina | 4 |
| Ácido glutâmico | 6 |
| Glutamina | 2 |
| Histidina | 2 |
| Isoleucina | 2 |
| Leucina | 6 |
| Lisina | 5 |
| Metionina | 2 |
| Prolina | 5 |
| Serina | 3 |
| Tirosina | 1 |
| Treonina | 3 |
| Triptofano | 2 |
| Valina | 3 |
| Selenocisteína | 0 |



Anexo C. Gráfico *dotplot* gerado pelo programa SyMAP que fez uso do programa PROmer para gerar as comparações ao nível de aminoácidos e construir blocos sintênicos pelo algoritmo de sintenia do SyMAP. Os blocos sintênicos estão representados pelas caixas azuis; dentro dessas, os blocos sintênicos são representados por linhas diagonais, as que vão do inferior esquerdo ao superior direito são inversões com relação ao sentido de cada *scaffold*.