

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

Anderson Carvalho Morais

Genômica comparativa de tripanossomatídeos portadores de endossimbiontes e sua  
implicação no genoma de *Angomonas ambiguus*

São Paulo  
2018

Anderson Carvalho Morais

Genômica comparativa de tripanossomatídeos portadores de endossimbiontes e sua  
implicação no genoma de *Angomonas ambiguus*

Versão Original

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

Área de concentração: Bioinformática

Orientador: Prof.<sup>o</sup> Dr.<sup>o</sup> João Marcelo Pereira Alves

São Paulo  
2018

Nome: MORAIS, Anderson Carvalho

Título: Genômica comparativa de tripanossomatídeos portadores de endossimbiontes e sua implicação no genoma de *Angomonas ambiguus*

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

Agradeço,

Ao Prof.<sup>o</sup> Dr.<sup>o</sup> João Marcelo Pereira Alves, o Jota, pela orientação no mestrado, pela oportunidade de desenvolver esta pesquisa, por todos os ensinamentos em bioinformática, tanto biologia quanto em computação, pela paciência em lidar com meus problemas, e por ser um ótimo profissional e melhor ainda como pessoa.

À Universidade de São Paulo, às agências de fomento à pesquisa FAPESP e FAPEMA pelo apoio financeiro.

Ao Dr.<sup>o</sup> André Martins, por toda a ajuda, pelas soluções de problemas, até os mais triviais, aos *scripts* que fez rodar, e principalmente pela amizade.

Aos amigos dos laboratórios 107 e 109, Bruno, Carla, Carmen, Flávia, Gabriel, Herakles, Juliana, Luciana, Lyslaine, Marta Campaner, Oneida, Priscila, Tânia, Tarcila, em especial à Paola, por me receberem tão bem, por sempre me apoiarem mesmo quando eu queria desistir, e pela amizade e bons momentos. Aos amigos de laboratório, Hugo, Juliano, Thaís, Ana Clara, pelo companheirismo. Amigos do departamento de parasitologia e microbiologia, Marcell, Richard, Irina e Liliana, por todos os bons momentos juntos.

À minha família, em especial minha mãe, Rosa, por dedicar a vida a cuidar da sua família com amor incondicional, por me incentivar nos estudos desde muito pequeno, por nunca deixar desistir. Aos meus irmãos Aline e Adriano, pelo amor e carinho. Ao meu cunhado Carlos Anderson por todo apoio, e para meus sobrinhos Pérola e Benjamim, por trazerem alegria à família. A toda minha grande família, incluindo tios, primos, avós, obrigado pelo apoio. A Alexandre, por estar sempre

comigo, cuidando de mim, pelo apoio e por ser uma inspiração.

À minha grande amiga Ananda Martins, por me abrigar nos meus primeiros meses em São Paulo, pelos milhões de conselhos, por todos os sorrisos que me fez ter, por ser você. Ao meu outro grande amigo Marcos Alencar, Marquinhos, por ter se tornado um irmão para mim, sempre estando ao meu lado nos momentos mais difíceis e proporcionado momentos muito felizes. Aos meus amigos Leonardo, Fernanda, Milena, Rafaella, Laís, Amanda, Elias, Bruno, Ana Paula, Thaysa, Agostinho, Cinthia, Jair, Raissa, Thayza, Camila por serem os melhores amigos.

**Só tenho a agradecer todos vocês, muito obrigado!**

“A ciência traz ao homem a incerteza de uma certeza”  
Adan Fernandes da Silva

## RESUMO

MORAIS, Anderson Carvalho. **Genômica comparativa de tripanossomatídeos portadores de endossimbiontes e sua implicação no genoma de *Angomonas ambiguus***. 2018, XX f. Dissertação (Mestrado em Bioinformática) – Programa Interunidades em Bioinformática, Universidade de São Paulo, São Paulo, 2018.

Tripanossomatídeos hospedeiros de endossimbiontes (SHT) possuem características únicas, em aspectos morfológicos, bioquímicos e moleculares, quando comparados a tripanossomatídeos regulares. Dentre as sete espécies conhecidas atualmente, só *Angomonas ambiguus* (Filo Euglenozoa; Família Trypanosomatidae) parece apresentar uma incongruência evolutiva entre os genomas do hospedeiro (núcleo) e seu endossimbionte (TPE), uma  $\beta$ -proteobactéria. Para entender esta incoerência, dentre outros aspectos genômicos e evolutivos, este trabalho iniciou uma análise comparativa entre os genomas de seis espécies de SHTs e seus respectivos TPEs, através de ferramentas de bioinformática. Com os genomas sequenciados e montados, foram preditos e anotados os genes de rRNAs, tRNAs e codificadores de proteínas; identificados os genes ortólogos, com objetivo de realizar uma comparação do conteúdo gênico, e também para realizar a comparação posterior por meio de inferência filogenética por máxima verossimilhança entre todos os grupos ortólogos encontrados entre *A. ambiguus*, *A. deanei*, *A. desouzai*, *Strigomonas Galati*, *S. oncopelti* e *S. culicis*, e entre os respectivos endossimbiontes. Foram feitas também análises filogenômicas através da construção de superárvores. O genoma do endossimbionte é bem reduzido, com perda no conteúdo GCe de genes ortólogos, como visto em os outros TPEs. A anotação funcional dos genes a partir das vias metabólicas do KEGG e utilizando o programa ASGARD mostrou que o TPE contribui fornecendo genes para completar muitas das vias de aminoácidos do SHT, como em alguns aminoácidos essenciais, tal qual prolina, arginina, valina, leucina, e outros não essenciais, comprovando a relação mutualística entre endossimbionte e hospedeiro, como é visto em outros trabalhos. Também, comprovou-se por filogenia multigênica que *A. ambiguus* está mais relacionada com *A. desouzai*, enquanto *Candidatus Kinetoplastibacterium crithidii* ‘amb’ está diretamente relacionado com o endossimbionte de *A. deanei*. No caso das espécies do gênero *Strigomonas*, cada flagelado possui sua própria espécie de endossimbionte, em perfeita coespeciação. O que provavelmente aconteceu foi a hibridização entre as espécies *A. desouzai* e *A. deanei*, permanecendo o núcleo da primeira e o endossimbionte da segunda espécie.

**Palavras-chave:** *Angomonas ambiguus*, endossimbionte, genoma, bioinformática.

## ABSTRACT

MORAIS, Anderson Carvalho. **Comparative genomics of trypanosomatids with endosymbionts and their implication in the genome of *Angomonas ambiguus***. 2018, XX f. Dissertação (Mestrado em Bioinformática) – Programa Interunidades em Bioinformática, Universidade de São Paulo, São Paulo, 2018.

Symbiont-harboring trypanosomatids (SHT) have unique characteristics in morphological, biochemical and molecular aspects when compared to regular trypanosomatids. Among the seven known species, *Angomonas ambiguus* (Filo Euglenozoa, Family Trypanosomatidae) seems to present an evolutionary incongruence between the genomes of the host (nucleus) and its  $\beta$ -proteobacterial endosymbiont (TPE). To understand this incoherence, among other genomic and evolutionary aspects, this work performed a comparative analysis between the genomes of six species of SHTs and their respective TPEs, using bioinformatics tools. With sequenced and assembled genomes, genes for rRNAs, tRNAs or encoding proteins were predicted and annotated; the orthologous genes were identified, aiming at a comparison of the gene content and to carry out a phylogenetic inference by maximum likelihood among all orthologous groups found in *A. ambiguus*, *A. deanei*, *A. desouzai*, *Strigomonas galati*, *S. oncopelti* and *S. culicis*, and the respective endosymbionts. Phylogenomic analyses were also performed through the construction of supertrees. The genome of the endosymbiont is well reduced, with loss in GC content and orthologous genes, as seen in other TPEs. Functional annotation of genes from KEGG metabolic pathways and using the ASGARD program have shown that the TPE contributes by providing genes to complete many of the SHT essential amino acid pathways such as proline, arginine, valine, leucine, as well as other nonessential amino acid pathways, proving the mutual relationship between endosymbiont and host, as seen in other studies. Analysis of multigenic phylogenies suggested that *A. ambiguus* is more related to *A. desouzai*, whereas *Candidatus Kinetoplastibacterium crithidii* 'amb' is directly related to the *A. deanei* endosymbiont. In the case of species from the genus *Strigomonas*, each flagellate has its own species of endosymbionts, in perfect coespeciation. What probably happened was the hybridization between the species *A. desouzai* and *A. deanei*, remaining the nucleus of the first and the endosymbiont of the second species.

**Key-words:** *Angomonas ambiguus*, endosymbiont, genome, bioinformatics

## LISTA DE FIGURAS

<p>Fig. 1: Micrografia eletrônica de parte do kDNA de <i>Angomonas deanei</i>. Os pequenos <i>loops</i> são os minicírculos, a seta indica parte do maxicírculos.....</p>	19
<p>Fig. 2: Microscopia de luz de tripanossomatídeos do gênero <i>Angomonas</i> e <i>Strigomonas</i> portadores de endossimbiontes. 1-2, <i>A. deanei</i>. 3-4, <i>S. oncopelti</i> e não portadores de endossimbiontes. Setas apontam para formas opistomorfas. Escala = 10µm.....</p>	22
<p>Fig. 3: Representação esquemática das relações evolutivas dos SHTs do gênero <i>Angomonas</i> e <i>Strigomonas</i> (lado esquerdo), e dos TPEs (lado direito). No gênero <i>Strigomonas</i> vemos um padrão entre os dois cladogramas, ao passo que <i>A. ambiguus</i> está relacionada a <i>A. desouzai</i>, mas seu endossimbionte está mais próximo do endossimbionte de <i>A. deanei</i>.....</p>	26
<p>Fig 4. Fluxo de etapas desenvolvidas neste projeto.....</p>	42
<p>Fig. 5: Representação dos grupos ortólogos no genoma das seis espécies de endossimbiontes <i>Candidatus</i> Kinetoplastibacterium em ordem de ocorrência após o gene <i>dnaA</i>. Caixas na mesma coluna representam genes ortólogos identificados pelo OrthoMCL. Caixas azuis representam ortólogos que estão presentes em mais de um organismo; verdes, ortólogos presentes em único organismo (<i>singleton</i>); vermelhos, pseudogenes; brancos, ortólogos ausentes na espécie. Abreviaturas: CKcri ‘a’ denota <i>Ca. K. crithidii</i> ‘amb’; CKcri, <i>Ca. K. crithidii</i>; CKdes, <i>Ca. K. desouzai</i>; CKbla, <i>Ca. K. blastocrithidii</i>; CKgal, <i>Ca. K. galatii</i>; e CKonc, <i>Ca. K. oncopeltii</i>.....</p>	47
<p>Fig. 6: Comparação linear entre os genomas dos seis endossimbiontes na ordem <i>Ca. K. cri</i> ‘amb’, <i>Ca. K. cri</i>, <i>Ca. K. des</i>, <i>Ca. K. blas</i>, <i>Ca. K. gal</i>, <i>Ca. K. onc</i>.....</p>	48
<p>Fig. 7: Categorização funcional de proteínas nos genomas de três endossimbiontes do gênero <i>Angomonas</i>.....</p>	50
<p>Fig. 8: Genes únicos de cada espécie de SHT do gênero <i>Angomonas</i> categorizados pelo produto proteico no eggNOG.....</p>	51

Fig. 9: Mapa metabólico disponível pelo KEGG e, genes identificados pelo Asgard, diferenciando aqueles presentes e ausentes na via da arginina e prolina. Cor azul representa proteínas de <i>A. ambiguus</i> e cor vermelha, proteínas de <i>Ca. K. crithidii</i> ‘amb’.....	54
Fig. 10: Mapa metabólico disponível pelo KEGG e genes identificados pelo Asgard, diferenciando aqueles presentes e ausentes na via da biossíntese da valina, leucina e isoleucina. Cor azul representa proteínas de <i>A. ambiguus</i> e cor vermelha, proteínas de <i>Ca. K. crithidii</i> ‘amb’.....	56
Fig. 11: Mapa metabólico disponível pelo KEGG e genes identificados pelo Asgard, diferenciando aqueles presentes e ausentes na via da biossíntese da histidina. Cor azul representa proteínas de <i>A. ambiguus</i> e cor vermelha, proteínas de <i>Ca. K. crithidii</i> ‘amb’.....	57
Fig. 12: Mapa metabólico disponível pelo KEGG e genes identificados pelo Asgard, diferenciando aqueles presentes e ausentes na via biossintética da lisina. Cor azul representa proteínas de <i>A. ambiguus</i> e cor vermelha, proteínas de <i>Ca. K. crithidii</i> ‘amb’.....	58
Fig. 13: Diagrama de Venn exibindo grupos ortólogos compartilhados entres as seis espécies de tripanossomatídeos (SHT).....	60
Fig. 14: Diagrama de Venn exibindo grupos ortólogos compartilhados entres as seis espécies de endossimbiontes (TPE).....	60
Fig. 15. Árvore filogenética construída a partir das 4.428 árvores resultantes da análise de comparação de ortólogos dos tripanossomatídeos por meio da abordagem superárvore. Valores de <i>bootstrap</i> presentes em cada nó.....	63
Fig 16 Árvore filogenética construída a partir das 641 árvores resultantes da análise de comparação de ortólogos dos endossimbiontes por meio da abordagem superárvore. Valores de <i>bootstrap</i> presentes em cada nó.....	63

## LISTA DE TABELAS

Tabela 1: Dados comparativos dos genomas dos SHTs.....	43
Tabela 2: Dados comparativos dos genomas dos TPEs.....	43
Tabela 3. Genes encontrados no maxicírculo de <i>A. ambiguus</i> a partir da ferramenta <i>blast</i> contra banco de dados NR e espécies identificadas.....	52
Tabela 4: N <sup>o</sup> de árvores em que cada espécie aparece em um mesmo clado com <i>Angomonas ambiguus</i> .....	61
Tabela 5: N <sup>o</sup> de árvores em que cada espécie aparece em um mesmo clado com <i>Candidatus Kinetoplastibacterium crithidii</i> ‘amb’.....	61

## SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 FILO EUGLENOZOA.....	17
1.1.1 FAMÍLIA TRYPANOSOMATIDAE.....	19
1.1.2 TRIPANOSSOMATÍDEOS PORTADORES DE ENDOSSIMBIONTES	21
1.1.3 CLASSIFICAÇÃO TAXONÔMICA.....	24
1.2 BIOINFORMÁTICA.....	27
1.2.1 SEQUENCIAMENTO DE GENOMAS.....	30
1.2.2 MONTAGEM.....	32
1.2.3 PREDIÇÃO DE GENES.....	33
1.2.4 ANOTAÇÃO.....	34
2 OBJETIVOS.....	36
3 MATERIAIS E MÉTODOS.....	37
4 RESULTADOS E DISCUSSÃO.....	43
4.1 CARACTERIZAÇÃO GERAL DOS GENOMAS.....	39
4.2 ANÁLISE DO CONTEÚDO GÊNICO.....	45
4.3 ANOTAÇÃO.....	49
4.3.1 KINETOPLASTO.....	52
4.3.2 VIAS METABÓLICAS.....	53
4.4 IDENTIFICAÇÃO DE GENES ORTÓLOGOS PARA ANÁLISE FILOGENÉTICA..	59
4.5 FILOGENÔMICA.....	62
5 CONCLUSÃO.....	66
REFERÊNCIAS.....	67

## 1 INTRODUÇÃO

O termo simbiose (do grego *sym*, junto, e *bios*, vida) foi empregado pela primeira vez por Heinrich Anton de Bary em 1879 para designar a associação entre diferentes organismos coexistindo e/ou, até mesmo, originando uma nova espécie, após a descrição realizada pelo botânico Simon Schwendener (1867) de líquens, união de um fungo e um organismo fotossintetizante – alga ou cianobactéria (HONEGGER, 2000). Hoje sabemos que a simbiose aparece em espécies de todos os ecossistemas, tanto terrestres quanto aquáticos, desempenhando um papel fundamental no surgimento das atuais formas de vida e na geração da diversidade biológica.

Normalmente as relações entre diferentes espécies são classificadas como comensalismo, quando uma das espécies é beneficiada, sem causar danos à outra (também conhecida como cooperação ou protocooperação); parasitismo, tipo de relação na qual uma das espécies se beneficia ao custo de outra; e mutualismo, interação na qual ambos os organismos se beneficiam (D H BOUCHER; S JAMES; KEELER, 1982). Chamamos de endossimbionte um organismo que vive dentro do corpo ou célula de outro organismo.

O evento de endossimbiose talvez mais admiravelmente exemplificável foi o que levou à evolução das organelas eucariontes – mitocôndrias e cloroplastos, oriundos de  $\alpha$ -proteobactéria e cianobactéria, respectivamente (MORAN, 2006). A hipótese para origem dos grupos eucariotos foi proposta primeiramente por Lynn Margulis no seu ensaio “Symbiosis in cell evolution” (1981), o qual se embasou em diversos trabalhos publicados anteriormente (SCHIMPER, 1883; MERESCHKOWSKY, 1905; WALLIN, 1925).

Passados mais de 30 anos, vemos cada vez mais como estudos, principalmente os moleculares, aumentaram nosso conhecimento sobre a importância e evolução, origem e função da endossimbiose em grupos animais, vegetais e de microrganismos. Ela pode alterar diversos efeitos ecológicos e evolutivos sobre o hospedeiro, influenciando desde aspectos da interação ecológica (BAKER 2003), nutrição (FIORE et al., 2010), defesa (MUSCATINE; MCNEIL, 1989), reprodução (BAZINET, 2004), até com consequências sobre as estruturas populacionais, como isolamento reprodutivo e especiação (KUTSCHERA; NIKLAS, 2005), dependendo de se a interação “hóspede-hospedeiro” evoluir para mutualismo, parasitismo ou relações intermediárias. No caso da primeira situação, normalmente o endossimbionte propicia algumas vantagens ao hospedeiro, o qual oferece um ambiente protegido e cheio de nutrientes.

Quando tratamos de endossimbiose, as bactérias são os seres normalmente mais bem estudados e relatados, sendo conhecidos casos de endossimbionte bacteriano em diversos grupos, como insetos (MORAN; MCCUTCHEON; NAKABACHI, 2008), esponjas (TAMAS et al., 2002), moluscos (KUWAHARA et al., 2008; NEWTON et al., 2007), nematoides (TAYLOR; BANDI; HOERAUF, 2005), e o foco deste trabalho, tripanossomatídeos (DE SOUZA; MOTTA, 1999).

Este grupo de protozoários flagelados é exclusivamente composto por parasitas de plantas ou animais, predominantemente invertebrados, durante todo seu ciclo de vida. Uma pequena exceção são os tripanossomatídeos que possuem dois hospedeiros, um deles podendo ser seres humanos. Dentre os com modo de vida monoxênico, encontramos um grupo com sete espécies de tripanossomatídeos que carregam bactéria endossimbionte,  $\beta$ -proteobacteria, distribuídas nos gêneros *Strigomonas*, *Angomonas* (TEIXEIRA et al., 2011) e *Kentomonas* (VOTÝPKA et al., 2014).

A relação intrínseca entre os dois organismos através da simbiose faz com que o tripanossomatídeo e sua respectiva bactéria endossimbionte sejam mantidos juntos, visto que o endossimbionte é incapaz de viver fora de seu hospedeiro e, da mesma maneira, culturas de tripanossomatídeos tratadas com antibióticos (apossimbióticas) são incapazes de colonizar insetos (DE SOUZA; MOTTA, 1999). Quando tratamos de microrganismos, os estudos a respeito da endossimbiose foram, por muito tempo, dificultados devido à impossibilidade de cultivar a maioria deles, e ainda hoje não foi possível cultivar muitos procariotos e eucariotos unicelulares. Entretanto, com o advento do sequenciamento genômico e da bioinformática, viabilizou-se uma análise detalhada das interações desses organismos.

Estudos envolvendo bioinformática mostram como ocorre a colaboração entre os SHTs e TPEs (ALVES et al., 2013; KOSTYGOV et al., 2017), através de uma abordagem comparativa dos genomas entre as espécies é possível verificar, além do conteúdo genético – o que há igual ou diferente–, inferir as relações evolutivas entre as mesmas, analisar o conjunto proteico compartilhado, analisar quais vias metabólicas estão associadas entre SHT e TPE, verificar o grau de sintonia entre os genomas, dentre uma série de investigações que dados oriundos de sequenciamento podem proporcionar.

A análise *in silico* de genomas surge tentando solucionar uma série de questionamentos que abordagens tradicionais não conseguem, principalmente quando tratamos de microrganismos que são pouco estudados por não serem de interesse médico ou veterinário. Estabelecer um método eficaz de análise de sequenciamento e bioinformática, baseado nas mais recentes tecnologias, é de grande importância para a ampliação dos projetos envolvendo o estudo desse grupo, que é encontrado parasitando uma série de hospedeiros de diversos grupos animais e vegetais.

Ainda, a bioinformática contribui imensamente na análise e na descoberta de novos dados para estudo do genoma de diversos organismos, incluindo grupos pouco estudados nos grandes projetos de pesquisa. A grande quantidade de dados que o sequenciamento de um genoma produz é enorme; assim, a análise dos mesmos deve ser adequada de acordo com as atuais ferramentas computacionais, para extrair o máximo de informações que esses dados podem trazer ao entendimento da evolução dos organismos e relações com outros grupos.

### 1.1 FILO EUGLENOZOA

O filo Euglenozoa (Eukaryota, Excavata) (CAVALIER-SMITH, 1981) engloba três grupos monofiléticos, euglenídeos, diplomonemídeos e cinetoplastídeos, consistindo de seres eucariotos, unicelulares e com uma grande variedade de modos de nutrição, incluindo predação, osmotrofia, parasitismo e fotoautotrofia (LEANDER, 2004). Como principais sinapomorfias do grupo temos um padrão de raiz flagelar com dois corpos basais e três arranjos de raízes microtubulares assimétricos; haste paraxial em um ou ambos os flagelos; e organelas extrusivas, tubulares e de paredes espessas, como tricocystos, mucocystos etc. (SIMPSON, 1997)

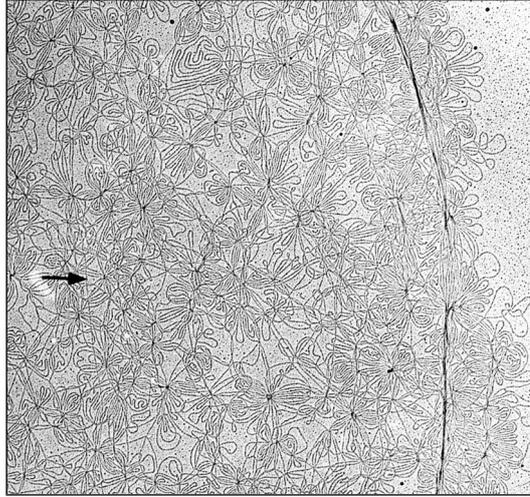
Os euglenídeos (classe Euglenoidea) são bem conhecidos devido a seus membros fotossintetizantes, como *Euglena*. Eles são agrupados por apresentarem um revestimento particular: um sistema superficial formado por quatro componentes principais – a membrana plasmática, um padrão de tiras proteicas sobrepostas, associadas à cisternas tubulares do retículo endoplasmático e microtúbulos (LEANDER; FARMER, 2000); a presença dessa estrutura dá um aspecto estriado à célula e propicia movimentos de ondulações ao organismo. Este

grupo evoluiu de um grupo de organismos fagotróficos que absorveram algas eucarióticas, o que levou, eventualmente, à aquisição do cloroplasto, através de uma endossimbiose secundária (GIBBS, 1981).

Os clados Euglenoidea e Kinetoplastea compartilham várias características morfológicas, incluindo flagelos com numerosas projeções finas semelhantes a pelos, haste paraxial heteromórfica e cristas mitocondriais discoidais (forma de pá) (KIVIC; WALNE, 1984). Enquanto Diplonemea possui uma grande mitocôndria com cristas achatadas e flagelos que, aparentemente, são lisos (MARANDE; LUKES; BURGER, 2005). Apesar das grandes diferenças entre os três grupos, o filo Euglenozoa é considerado monofilético baseado em análises filogenéticas moleculares, incluindo análise com gene SSU rRNA e genes codificadores de proteínas (BREGLIA; YUBUKI; LEANDER, 2013).

Os cinetoplastídeos (que incluem seres de vida livre, como os bodonídeos, e parasíticos, como *Trypanosoma* e *Leishmania*) apresentam como sinapomorfia a alteração do DNA mitocondrial em cinetoplasto ou kDNA, uma rede de moléculas de DNA circular associadas a proteínas, formando um corpúsculo denso dentro da mitocôndria (SIMPSON; LUKES; ROGER, 2002), localizado na base do flagelo. Nele existem os maxicírculos e os minicírculos (Fig. 1), sendo que os primeiros variam de 20 e 35 kb de comprimento, dependendo da espécie. É nele que está presente a maioria dos genes codificadores de proteínas mitocondriais, como rRNAs e subunidades do complexo respiratório (JENSEN; ENGLUND, 2012). Os minicírculos são mais variáveis, tanto em tamanho quanto em conteúdo; na família Trypanosomatidae, variam de 400 bp a 10 kb (GREIF et al., 2015). O papel deles é codificar uma pequena molécula de RNA identificada como RNA guia (gRNA), a qual direciona a edição dos transcritos vindos dos maxicírculos (adição/deleção de uridinas) (MADINA et al., 2014).

**Fig. 1:** Micrografia eletrônica de parte do kDNA de *Angomonas deanei*. Os pequenos *loops* são os minicírculos, a seta indica parte do maxicírculo.



Fonte: Englund et al., 2005.

Estudos moleculares baseados na análise genética do gene de RNA 18S separaram os cinetoplastídeos em dois grandes grupos, Prokinetoplastina – compreendendo a ordem Prokinetoplastida– e Metakinetoplastina, com quatro ordens: Eubodonida, Parabonida, Neobodonida e Trypanosomatida (que compreende a família Trypanosomatidae) (Moreira et al., 2004). Os bodonídeos, representados pelas três primeiras ordens, são geralmente espécies de vida livre, biflageladas, encontrados em ambiente aquático, e desempenham papel ecológico fundamental nesses ambientes, e é considerado o grupo mais ancestral aos tripanossomatídeos (D’AVILA-LEVY et al., 2009).

### 1.1.1 FAMÍLIA TRYPANOSOMATIDAE

As espécies classificadas na família Trypanosomatidae apresentam uma grande diversidade de organismos, com importância médica, veterinária e econômica. Todos são parasitas obrigatórios; e depois dos nematoides, os tripanossomatídeos são os eucariotos que apresentam a maior diversidade de hospedeiros, infectando animais invertebrados como artrópodes hematófagos ou

fitófagos, plantas, sanguessugas e todas as classes de vertebrados (VICKERMAN, 1994; SIMPSON; STEVENS; LUKEŠ, 2006; STEVENS, 2008; CAVALIER-SMITH, 2010).

Tripanossomatídeos heteroxênicos representam um dos grupos com maior interesse em pesquisa, no filo Euglenozoa, visto que há interesse médico, veterinário e econômico, que inclui gêneros *Trypanossoma*, *Leishmania*, *Endotrypanum* e *Porcisia*, parasitas de mamíferos, e *Phytomonas*, parasitas de plantas – completando seu ciclo de vida em insetos vetores (KAUFER et al., 2017). As espécies apresentando modo de vida monoxênico, que são as predominantes, incluindo 14 gêneros, são bons modelos para o estudo da diversidade parasítica de protistas, associação hospedeiro-parasita e evolução de eucariotos (MOTTA, 2010). O ciclo de vida dos tripanossomatídeos, tanto heteroxênicos quanto monoxênicos, é complexo, visto que apresentam diversas formas que variam de acordo com a espécie e com a fase do ciclo de vida, podendo ser tripomastigota, promastigota, amastigota, opimastigota, epimastigota e coanomastigota (DE SOUZA et al., 2002).

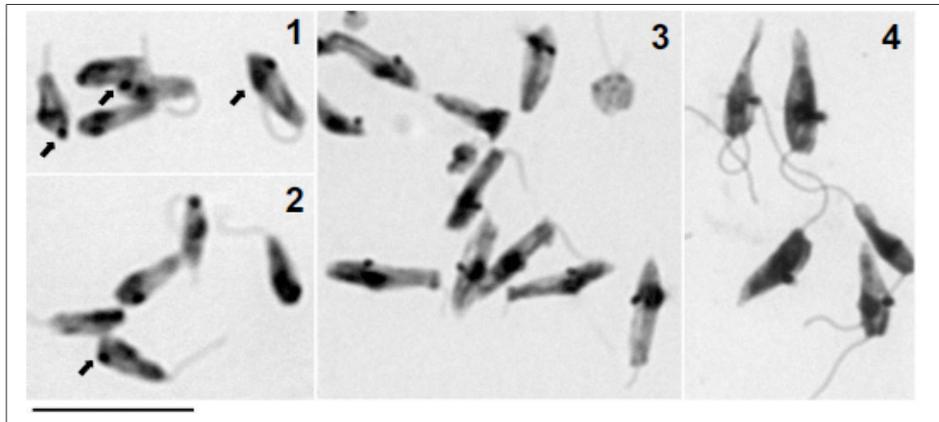
Acredita-se que o modo de vida parasítico tenha se originado a partir de um ancestral de vida livre, hoje relacionado ao grupo dos eubodonídeos; os primeiros grupos monoxênicos teriam surgido depois que insetos ingeriram essas espécies de vida livre, as quais se adaptaram ao intestino destes artrópodes (STEVENS, 2008). Com a capacidade de se alimentar de sangue, os insetos passaram a inocular os parasitas em vertebrados; os que se adaptaram começaram a circular entre os vertebrados e invertebrados, originando as espécies heteroxênicas (HAMILTON et al., 2004). Análises filogenéticas da família mostram que o parasitismo heteroxênico surgiu múltiplas vezes e em diferentes momentos na evolução do grupo (SIMPSON; STEVENS; LUKEŠ, 2006; STEVENS, 2008).

### 1.1.2 TRIPANOSSOMATÍDEOS PORTADORES DE ENDOSSIMBIONTES

Tem-se conhecimento de que sete espécies de tripanossomatídeos dos gêneros *Angomonas*, *Strigomonas* e *Kentomonas*, todos parasitas somente de insetos, são portadores de uma bactéria endossimbionte (CHANG, 1975; FIORINI et al., 1989; FARIA E SILVA et al., 1991; MOTTA et al., 1991; TEIXEIRA et al. 2011, VOTÝPKA et al., 2014), denominados SHT, do inglês *Symbiont-Harboring Trypanosomatids*. As bactérias por sua vez são denominadas *Candidatus Kinetoplastibacterium* ou TPE, do inglês *Trypanosomatid Proteobacterial Endosymbiont*. Pertencentes à classe  $\beta$ -proteobactéria, são intracitoplasmáticos obrigatórios, não sobrevivendo fora do hospedeiro, além de possuir uma dupla membrana separada por um espaço eletrolítico e dividem-se sincronicamente com seu hospedeiro (DE SOUZA; MOTTA, 1999), e há, normalmente, somente um TPE por SHT (MOTTA et al., 2010).

O primeiro endossimbionte descrito em um tripanossomatídeo foi caracterizado como um diplossomo (estrutura formada por dois centríolos), em *Strigomonas culicis* isolado de *Triatoma infestans* (NOVY; MACNEAL; TORREY, 1907). Posteriormente, em *Strigomonas oncopelti* uma estrutura similar foi denominada corpo bipolar, capaz de se auto reproduzir e composta por ribonucleoproteínas (NEWTON; HORNE, 1957). Logo, as mesmas estruturas foram reportadas em *Angomonas desouzai* e *Angomonas deanei* (espécies então classificadas no gênero *Crithidia*, ver abaixo). Após o tratamento por antibióticos de culturas com essas espécies, conclui-se que, na verdade, os diplossomos e corpos bipolares correspondiam à mesma estrutura, endossimbiontes (Fig. 2) (MUNDIM et al., 1974).

**Fig. 2:** Microscopia de luz de tripanossomatídeos do gênero *Angomonas* e *Strigomonas* portadores de endossimbiontes. 1-2, *A. deanei*. 3-4, *S. oncopelti* e não portadores de endossimbiontes. Setas apontam para para formas opistomorfas. Escala = 10 $\mu$ m.



Fonte: Borghesan, 2013.

Análises moleculares do DNA ribossômico mostraram que os TPEs de *S. culicis* e *A. deanei* compartilham alta identidade (DU; MASLOV; CHANG, 1994), o que sugere que tenham se originado do mesmo ancestral, uma  $\beta$ -proteobactéria (DU; MCLAUGHLIN; CHANG, 1994). Em conjunto, os dados moleculares sugerem que a endossimbiose em tripanossomatídeos tenha surgido somente uma vez, no período Cretáceo, no qual um hospedeiro ancestral adquiriu uma proteobactéria, sucedendo na coevolução de ambos e por fim também na coespeciação.

A presença de um endossimbionte induz mudanças morfológicas, bioquímicas e moleculares no hospedeiro tripanossomatídeo, quando tratados com antibióticos, tornando a cultura apossimbiótica, parte dessas características únicas são perdidas (ROITMAN; CAMARGO, 1985). Morfologicamente, o hospedeiro não possui a estrutura paraflagelar, localizado no flagelo; nem microtúbulos abaixo da membrana plasmática, onde as ramificações das mitocôndrias tocam; por último, as fibras do kDNA são mais dispersas, fazendo com que o mesmo seja mais frouxo, quando comparado ao kDNA de outros tripanossomatídeos (DE SOUZA; MOTTA, 1999).

Quando tratamos de aspectos bioquímicos, alguns trabalhos demonstram como os endossimbiontes contribuem com o metabolismo dos seus hospedeiros (CHANG, 1975; ROITMAN; CAMARGO, 1985; ALVES et al., 2013). Um dos primeiros trabalhos sobre como o TPE colabora nutricionalmente com o SHT, mostra com os hospedeiros não exigem adição, ao meio de cultura, de ornitina, arginina e citrulina, nutrientes intermediários do ciclo da ureia que são supridos pelo endossimbionte (GALINARI; CAMARGO, 1978).

Um aspecto particularmente significativo é a importante relação mutualística entre os dois organismos simbiotes, onde o tripanossomatídeo fornece um habitat propício para a sobrevivência de TPEs, enquanto estes, por sua vez, fornecem nutrientes essenciais para o metabolismo do seu hospedeiro (CHANG, 1975; ROITMAN; CAMARGO, 1985; ALVES et al., 2013). Por exemplo, estudos sobre requerimentos nutricionais mostraram que tripanossomatídeos hospedeiros de TPEs não exigem adição, ao meio de cultura, de ornitina, arginina e citrulina, nutrientes intermediários do ciclo da ureia que são supridos pelo endossimbionte (GALINARI; CAMARGO, 1978).

Outro importante nutriente fornecido pela bactéria endossimbionte de tripanossomatídeos é o grupamento heme (composto carregador de  $\text{Fe}^{2+}$  e fundamental para a maioria dos seres vivos). Os SHTs não requerem de hemina (uma fonte de heme) no meio de cultura, indicando a capacidade do TPE de sintetizar o heme. A via completa da síntese desse composto é encontrada na maioria dos organismos e é altamente conservada (MUNDIM; ROITMAN, 1977; MUNDIM et al., 1974; ALFIERI; CAMARGO, 1982; FARIA E SILVA et al., 1991). Estudos recentes baseados em análises de genomas completos mostraram a presença dos genes *gltX*, *CPOX*, *PPOX* e *FeCH* em espécies do gênero *Angomonas* e *Strigomonas*, que codificam algumas das enzimas da via biossintética do heme,

oriundos de transferência lateral (exceto *gltX*) de uma ou mais bactérias da classe Gammaproteobacteria; enquanto que para genomas de TPEs, apresentam os outros genes que complementam a via do heme (*hemA*, *GSA*, *ALAD*, *PBGD*, *UROS*, *UROD*, *hemN*, *FeCH*), os TPEs também codificam o *gltX* (ALVES et al., 2011).

Apesar dos diversos estudos envolvendo a família Trypanosomatidae, são poucos os que examinam o genoma total dos tripanossomatídeos e seus endossimbiontes. Alves e cols. (2013b) analisaram cinco genomas de TPEs, mostrando que a evolução genômica destas  $\beta$ -proteobactérias foi bem similar à de endossimbiontes de grupos bem distintos dos tripanossomatídeos, tais como amebas e insetos. Como principais aspectos revelados pelo estudo temos: diminuição do genoma total por perda de genes e retenção preferencial de genes envolvidos na colaboração entre endossimbionte e hospedeiro, com adicional pressão seletiva positiva em alguns casos. Outras características semelhantes a outros simbiontes incluem redução do conteúdo de GC, a ausência completa de rearranjo nos genomas e a perda de alguns genes relacionados à recombinação não homóloga.

### 1.1.3 CLASSIFICAÇÃO TAXONÔMICA

Devido à ausência de registros fósseis, a grande maioria das pesquisas relacionadas à taxonomia e filogenia de tripanossomatídeos tem sido feita com base em morfologia, ciclo de vida e preferência de hospedeiros. Entretanto, devido tanto a limitações nestes tipos de dados para a inferência filogenética quanto a características peculiares de tripanossomatídeos que albergam endossimbiontes, a utilização desses caracteres torna-se uma tarefa imprecisa. Devido a isso, estudos filogenéticos recentes têm utilizado diversos marcadores moleculares tentando desvendar e compreender a evolução e as relações filogenéticas do grupo

(HOLLAR; MASLOV, 1997; HUGHES; PIONTKIVSKA, 2003; YURCHENKO et al., 2006).

Estudos recentes, baseados em análises dos genomas completos dos simbiontes de tripanossomatídeos, mostraram a total sintonia dos genes das espécies de endossimbiontes e confirmaram que estes pertencem a classe  $\beta$ -proteobactéria, família Alcaligenaceae (MOTTA et al., 2013), e são mais proximamente relacionados com o gênero *Taylorella* (ALVES et al., 2013b).

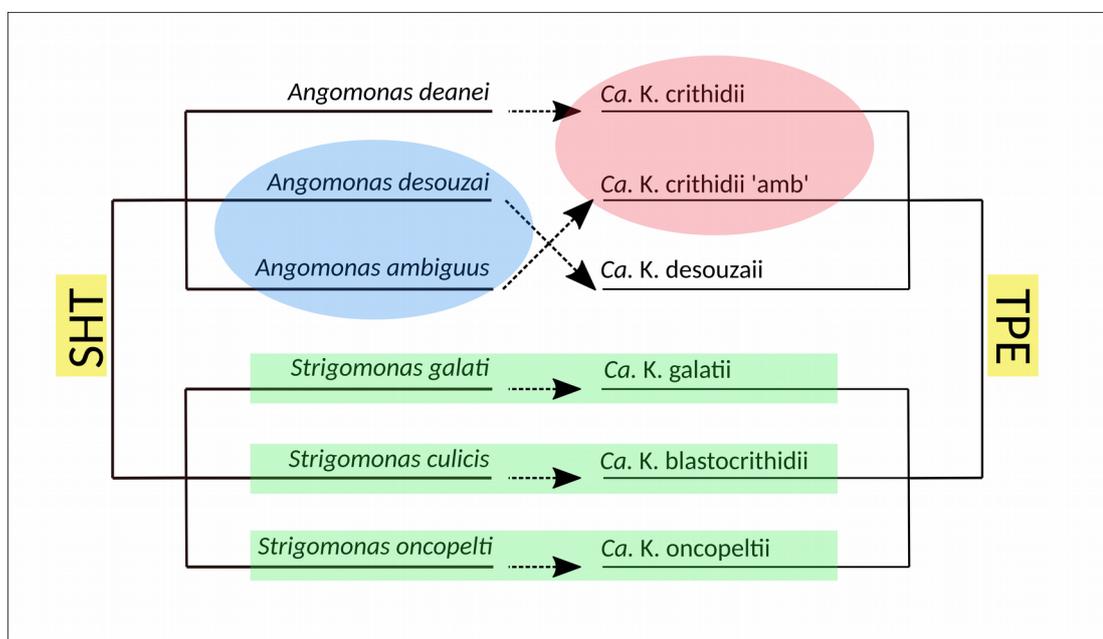
Em relação ao hospedeiro tripanossomatídeo, Souza e Corte-Real (1991) propuseram uma mudança na classificação de *Crithidia deanei* em um novo gênero, chamado de *Angomonas*, a partir de caracteres morfológicos. Já em 2000, Brandão e colaboradores, através de análises de caracteres morfológicos e kDNA, propuseram a classificação de *C. deanei* e *C. desouzai* no gênero *Angomonas*. Entretanto, essa classificação não foi adotada, devido, provavelmente, à falta de dados taxonômicos suportados por análises filogenéticas moleculares. Em 2011, Teixeira e colaboradores analisaram e estabeleceram, a partir de sequências da região V7-V8 da subunidade menor do RNA (SSU rRNA), além de critérios nutricionais e morfológicos, o gênero *Angomonas* e grupos relacionados. Como resultado, o gênero *Angomonas* foi dividido em dois subgrupos: *deanei* e *desouzai*. O primeiro grupo é composto por *Angomonas deanei* e mais seis parátipos, incluindo o isolado previamente denominado *Herpetomonas roitmani*. O segundo compreende *Angomonas desouzai*, dois parátipos e uma nova espécie, *Angomonas ambiguus* (TEIXEIRA et al., 2011).

Como dito anteriormente, acredita-se que a associação de tripanossomatídeos com bactérias endossimbiontes originou-se em um único momento no período Cretáceo (DU; MASLOV; CHANG, 1994). Dessa maneira, os grupos que apresentam endossimbionte são monofiléticos e bem distintos dos clados de

tripanossomatídeos que não o apresentam. Dentro dos clados, também há uma congruência entre as espécies: análises filogenéticas confirmam que os TPEs do gênero *Strigomonas* são bem relacionados entre si e suas árvores filogenéticas apresentam o mesmo padrão de ramificação daquele visto entre os SHTs.

Da mesma maneira, a filogenia dos TPEs do gênero *Angomonas* reflete a de seus respectivos SHTs, com exceção de *Angomonas ambiguus* (TEIXEIRA et al., 2011). O genoma nuclear de *A. ambiguus* parece ser filogeneticamente mais relacionado a *A. desouzai*, enquanto o endossimbionte parece ser mais proximamente relacionado a *A. deanei* (este cenário é exibido na Fig. 3). Por tal motivo, o endossimbionte de *A. ambiguus* foi classificado como *Candidatus Kinetoplastibacterium crithidii* (neste trabalho será usado denominado como *Candidatus Kinetoplastibacterium crithidii* 'amb') (TEIXEIRA et al., 2011).

**Fig. 3:** Representação esquemática das relações evolutivas dos SHTs do gênero *Angomonas* e *Strigomonas* (lado esquerdo), e dos TPEs (lado direito). No gênero *Strigomonas* vemos um padrão entre os dois cladogramas, ao passo que *A. ambiguus* está relacionada a *A. desouzai*, mas seu endossimbionte está mais próximo do endossimbionte de *A. deanei*.



Como mencionado acima, parece haver uma discordância entre a origem do genoma do tripanossomatídeo e a origem do genoma de seu endossimbionte na espécie *Angomonas ambiguus*. E recentemente, a filogenômica emergiu como um novo domínio da filogenia, tentando solucionar algumas questões que a filogenia molecular mais tradicional, que usa um ou dois genes, não consegue resolver. Esta abordagem apresenta como principal vantagem a menor probabilidade de erros de análise oriundos de fatores estocásticos, já que o uso de um grande conjunto de dados aumenta consideravelmente a consistência estatística de reconstruções e testes filogenéticos (GEE, 2003). Empregar a análise filogenômica no estudo de SHT e TPE, irá permitir definir graus de discordância filogenética, como o caso de *A. ambiguus* e seu endossimbionte e essencial para melhor entendimento da evolução dos mesmos.

## 1.2 BIOINFORMÁTICA

O avanço na tecnologia de sequenciamento, em especial com a segunda geração de sequenciadores, permitiu um aumento considerável no número de sequências genômicas disponíveis em bancos de dados, devido, principalmente, ao seu alto rendimento e à diminuição de seus custos (SHENDURE; JI, 2008). A possibilidade de sequenciar o genoma de um organismo viabilizou que diversas abordagens de estudos fossem desenvolvidas, permitindo obter, por exemplo, a estrutura e função dos genes, variedade genética, presença de elementos móveis, expressão diferencial, presença de genes adquiridos por transferência lateral e construção de mapas metabólicos (NIERMAN; EISEN; FRASER, 2000).

Sequências genômicas completas, a princípio, constituem uma fonte de dados singular, representando aquilo que é necessário para a formação de um organismo

em conjunto com os fatores epigenéticos. A análise do genoma de um único indivíduo, considerado isoladamente, não é capaz de revelar muito sobre si. Para tanto, o estudo comparativo entre genomas de espécies relacionadas pode, muitas vezes, prever as funções de genes, através da transferência de anotação funcional de proteínas a partir de organismos mais bem estudados para menos estudados (BROSCH et al., 2001). Isto torna genômica comparativa uma abordagem poderosa para a obtenção de uma melhor compreensão dos genomas. Além disso, ela tem auxiliado em aspectos fundamentais na genética, bioquímica e investigação das relações filogenéticas entre as espécies (GALPERIN; KOONIN, 1999; KONDRASHOV, 1999; FRASER et al., 2000; HUYNEN et al., 2005; ABBY; DAUBIN, 2007).

Apesar do atual uso de sequências moleculares para entender as relações entre grupos de espécies, desde a década de 1980 já usavam-se genes para a inferência da filogenia. Entretanto, a informação de um único gene era frequentemente insuficiente para obter um suporte estatisticamente relevante para a filogenia. Com o desenvolvimento das tecnologias de sequenciamento e a obtenção de uma grande quantidade de dados genéticos, tal problema pode ser amenizado devido a efeitos na amostragem, expandindo o número de caracteres e táxons usados na análise filogenética (DAUBIN; GOUY; PERRIÈRE, 2002). Como consequência, foi possível revelar as relações entre grupos e espécies que antes, usando caracteres morfológicos ou somente poucos genes, não era possível de determinar, principalmente quando tratamos de grupos de microrganismos.

No âmbito da biologia molecular, as relações evolutivas entre as espécies são deduzidas através da comparação de sequências, tentando encontrar uma indicação indireta de homologia entre os genes. Genes homólogos são aqueles que descendem de um ancestral em comum; partindo desse conceito, genes ortólogos são homólogos

que divergiram entre si após um evento de especiação (FITCH, 1970; 2000). Já genes parálogos surgiram após evento de duplicação gênica, e *inparalogs* são parálogos que se originaram a partir de duplicações linhagem-específicas após o evento de duplicação; por sua vez, *outparalogs* são resultantes de duplicações que precederam um dado evento de especiação (SONNHAMMER; KOONIN, 2002). Esses conceitos são importantes pois um dos métodos utilizados para comparação de genomas e posterior inferência filogenética é baseado na ausência ou presença de famílias gênicas observadas nos genomas, um deles através de genes ortólogos.

Outro modo de comparação de genomas se dá através de investigações de sintenia e mudança de organização estrutural de genes. Sintenia refere-se à compreensão de como a ordem dos genes em genomas foi se modificando ao longo do tempo, caso venham a ser alterados. Análise de sintenia permite examinar atentamente as forças seletivas e de mutações que atuam sobre a estrutura do genoma (GHEDIN et al., 2004). Entender a ordem de genes e a estrutura do genoma em diferentes espécies pode funcionar como marcador para o tempo de divergência entre elas.

A filogenômica surge então no sentido de interseccionar os campos da filogenia e genômica (EISEN; FRASER, 2003), e dessa forma visa a integração de dados genômicos, sejam estes na forma de sequências de nucleotídeos/aminoácidos, propriedades de sequências (como conteúdo GC, entropia e classificação funcional), propriedades genômicas (sintenia, rearranjos, abundância relativa de domínios) ou dados vindos de análises de microarranjos, com arcabouço evolutivo fornecido pela filogenética (SJÖLANDER, 2004).

### 1.2.1 SEQUENCIAMENTO DE GENOMAS

O sequenciamento genômico é a técnica que permite identificar a sequência de nucleotídeos de uma molécula de DNA ou RNA, na ordem correta, com o objetivo final de determinar a informação genética presente na estrutura. Com o passar do tempo, a tecnologia de sequenciamento foi se modificando e aprimorando; um dos primeiros métodos utilizados foi por hidrólise ou degradação química, conhecido como Maxam–Gilbert (MAXAM; GILBERT, 1977). Logo depois veio o sequenciamento por terminação de cadeia com dideoxynucleotídeos de Sanger, que é utilizado ainda hoje. Em 1986 foi lançado o primeiro sequenciador automático e, em 1998, o primeiro sequenciador por eletroforese capilar. Com a automatização do processo foi possível realizar grandes projetos de sequenciamento.

A segunda geração da tecnologia de sequenciamento, conhecida como NGS (*Next Generation Sequencing*), veio com uma proposta de acelerar e diminuir os custos do processo de sequenciamento. As plataformas NGS, apesar de possuírem tecnologias particulares, se baseiam no processo massivo de fragmentos de DNA, que podem ler milhões até bilhões de fragmentos ao mesmo tempo, ao contrário de tecnologias anteriores. Em 2004, com o lançamento do sequenciador da Roche 454, e logo depois com o lançamento das plataformas SOLEXA e Illumina em 2005, houve a primeira queda acentuada nos custos por megabase sequenciada. Apesar disso, o processo da reação e de equipamentos ainda custavam caro (HEATHER; CHAIN, 2016).

As tecnologias utilizadas pelas plataformas de sequenciamento de segunda geração produzem uma quantidade de leituras muito maior se comparadas ao Sanger, mas o tamanho médio delas é menor. O método de pirosequenciamento com detecção de pirofosfato, utilizado pelo Roche 454 produz leituras com média

de 500 bp, uma atualização em 2008, com a plataforma 454 GS FLX Titanium passou a produzir leituras de 700 bp, apesar da produção de longas leituras, esse tipo de tecnologia possui uma baixa acurácia para representar regiões homopoliméricas (ZHANG et al., 2011).

Outras tecnologias NGS, como o ABI SOLiD (sequenciamento por ligação), em 2006 produzia leituras de tamanho 25-35 bp, atualmente, o tamanho varia de acordo com a biblioteca utilizada (Mate-paired com 2x60 pb, Paired-end com 75x35 pb e Fragment com 75 pb); IonTorrent (sequenciamento por ionização), na plataforma atual consegue produzir leituras de até 200 bp; Illumina (sequenciamento por síntese), dependendo do modelo (MiSeq, NextSeq 500 e HiSeq) pode variar de 35 a 250 bp (REUTER; SPACEK; SNYDER, 2015).

Os sequenciadores de segunda geração utilizam estratégias diferentes, mas possuem etapas em comum:

- 1 – Preparo da amostra: o DNA é fragmentado de maneira aleatória; adaptadores são incorporados aos *templates* (fragmentos). A criação de bibliotecas de DNA pode ser feita de duas maneiras: paired-end, que sequencia as duas extremidades dos *templates* com tamanhos entre 200 a 500 nucleotídeos, ou *mate-pair*, sequenciando fragmentos de tamanhos maiores, entre 2000 a 5000 nucleotídeos. Nesse caso são gerados *templates* maiores, os adaptadores são ligados e o fragmento circularizado; é realizada uma digestão enzimática que gera um fragmento com duas pontas do *template* e com o adaptador na região interna. Essa forma de preparo, apesar de mais etapas, permite a detecção de variações estruturais e geração de *scaffolds*.

- 2 – Amplificação: tem como objetivo gerar milhares de cópias de cada fragmento de DNA produzido na etapa anterior, de modo a aumentar a fonte do

sinal luminoso (ou, no caso do IonTorrent, aumento do sinal iônico) que será detectado pelo sequenciador.

3 – Sequenciamento: nesta etapa um aparelho, o sequenciador, executa uma série de reações químicas, que geram sinais que são detectados e determinam a sequência do *template*.

Atualmente estamos na terceira geração de sequenciadores, lançada em 2010. A plataforma da Pacific Biosciences usa a tecnologia *single-molecule real-time* (SMRT) ou método de síntese em tempo real, que produz leituras longas, chegando a 60 kbp com precisão de 99,99% (se utilizada alta cobertura), entretanto ainda é uma tecnologia com custos bem altos. A tecnologia desenvolvida pela Oxford Nanopore Technologies, lançada nos últimos anos, utiliza-se do método de nanoporo onde são incorporados poros de proteína em uma membrana de polímero sintético dentro de um microchip. A cadeia de DNA a ser sequenciado passa através do poro e, quando cada nucleotídeo atravessa, é registrado como um sinal elétrico diferente detectado pela máquina (REUTER; SPACEK; SNYDER, 2015).

### 1.2.2 MONTAGEM

Ao sequenciar um genoma completamente, milhares de fragmentos são gerados, assim a montagem consiste em um conjunto de procedimentos que buscam organizar essas sequências curtas em um espaço linear, com o intuito de representar verdadeiramente o DNA da espécie em estudo. Primeiramente existe uma etapa de análise de qualidade, em que bases de baixa qualidade são removidas. Ela é verificada por um *score*, chamado de Phred, indicando o quanto o sequenciamento foi confiável para cada uma das bases; quanto maior o seu valor, maior será a confiança de que determinada base foi atribuída de forma correta em uma leitura.

Após a análise de qualidade, as leituras podem ser usadas para a montagem, onde são alinhadas umas com as outros, tentando identificar as regiões que são sobrepostas (SCHLEBUSCH; ILLING, 2012), formando assim *contigs*. Alta cobertura garante maior precisão no alinhamento das leituras e reduz a taxa de erros nas montagens (MILLER et al., 2010; SCHATZ et al., 2010).

Um dos parâmetros para calcular a qualidade da montagem é analisando o N50, que mostra o comprimento do menor *contig* a partir do qual o somatório de todos os *contigs* (do maior para o menor) representa a metade do tamanho do genoma, revelando o quanto de um genoma é coberto por *contigs* grandes (SCHATZ; DELCHER; SALZBERG, 2010).

### 1.2.3 PREDIÇÃO DE GENES

Uma das principais etapas em projetos de genoma é a identificação de genes codificadores de proteínas, mas também identificar outros elementos funcionais do DNA, como genes de RNA e regiões regulatórias. Nos últimos anos houve um crescente desenvolvimento de técnicas e programas que pudessem melhorar cada vez mais a forma como as sequências são preditas, em razão da grande quantidade de dados gerados por uma única corrida, e do fato de que são sequenciados genomas, principalmente de procariotos, a toda hora. Assim, realizar a tarefa de predição bem e de forma automatizada é essencial para bom andamento de qualquer projeto genoma (GOEL; SINGH; ASERI, 2013).

A predição de genes em eucariotos é uma etapa bastante complexa, dada a organização dos mesmos em éxons e introns; aqui, no entanto, será tratado da predição em procariotos, visto que os genes dos tripanossomatídeos estudados se comportam como de procariotos, por não possuírem íntrons (somente quatro nos

genomas previamente caracterizados). O primeiro passo para a análise do genoma procarioto é a identificação de todos os genes, visto que eles representam 90% do genoma total. Particularmente, o problema se dá em identificar ORFs (*Open reading frame* ou quadro aberto de leitura) verdadeiras, já que é comum encontrar ORFs parcial ou completamente sobrepostas, e localizar sinais regulatórios onde existe um promotor ou um terminador (GALPERIN; KOONIN, 1998).

As duas principais abordagens de predição são extrínseca e intrínseca. Na primeira, realiza-se a predição por comparação, realizando-se um BLAST (ALTSCHUL et al., 1990) com sequências de organismos próximos, o que pode ser bastante efetivo. Na abordagem intrínseca, o preditor usa aspectos da própria sequência para identificar os genes, um preditor que usa a segunda abordagem é conhecido como *ab initio*.

Existem diversos programas para predição de genes *ab initio*, como GeneMark (BESEMER; BORODOVSKY, 2005), que usa o modelo de cadeias de Markov (HMM) para identificar regiões codificadoras, e GLIMMER (SALZBERG et al., 1998), que usa modelo de Markov interpolado (MMI). O MMI é, em princípio, mais poderoso do que as cadeias de Markov, e alguns experimentos computacionais mostram resultados mais precisos quando usados para encontrar genes em DNA bacteriano.

#### 1.2.4 ANOTAÇÃO

A obtenção das sequências de DNA não fornece muitas informações biológicas de modo imediato e representa apenas o primeiro passo para a compreensão do genoma. O entendimento dos mecanismos envolvidos no funcionamento do genoma depende da precisão dos processos de anotação. Anotar

uma sequência nada mais é que determinar um significado biológico para ela (STEIN, 2001).

A anotação, atualmente, é realizada basicamente por comparação, com sequências depositadas em diversos bancos de dados, através principalmente do BLAST (*Basic Local Alignment Search Tool*), e suas variantes. Os principais bancos de dados são NR ( proteínas não-redundantes) , PFAM (banco de domínios proteicos), INTERPRO (integra famílias, domínios e sítios proteicos), CDD (banco de domínios conservados), KEGG (banco de vias metabólicas).

## 2 OBJETIVOS

### Objetivo Geral

O objetivo deste trabalho é realizar uma análise em larga escala do genoma de *Angomonas ambiguus* TCC2435 e seu endossimbionte através da abordagem comparativa com outros tripanossomatídeos e endossimbiontes.

### Objetivos Específicos

1. Caracterizar o genoma *A. ambiguus* (SHT) e *Candidatus Kinetoplastibacterium crithidii* 'amb' (TPE);
2. Realizar a anotação genômica do SHT e TPE;
3. Comparar o genoma de *A. ambiguus* com aqueles de outros tripanossomatídeos previamente sequenciados, em especial dos gêneros *Angomonas* e *Strigomonas*.
4. Comparar o genoma do TPE de *A. ambiguus* com os genomas de outros endossimbiontes de tripanossomatídeos;
5. Confirmar, por filogenia multigênica (filogenômica), a posição filogenética de *A. ambiguus* e seu TPE em relação aos outros SHTs e TPEs já sequenciados.

### 3 MATERIAIS E MÉTODOS

#### 3.1 SEQUÊNCIAS E MONTAGEM DO GENOMA

O departamento de Parasitologia do Instituto de Ciências Biomédicas (USP) junto com Department of Microbiology and Immunology, Virginia Commonwealth University, EUA desenvolveram um grande projeto denominado *Assembling the Tree of Life – Phylum Euglenozoa* (AToLE), com o intuito de estudar especialmente organismos não-parasitas, ou parasitas sem interesse médico humano. Este grande projeto fornece subsídios para desenvolver uma série de pesquisas envolvendo o Filo Euglenozoa (Euglenida, Diplonemida e Kinetoplastida), estando o Prof.<sup>o</sup> Dr.<sup>o</sup> João Marcelo Pereira Alves encarregado de estudar os genomas e transcriptomas de diversos organismos da classe Kinetoplastida, além do desenvolvimento de ferramentas de bioinformática. Muitos dos dados genômicos já disponíveis a serem utilizados neste trabalho provêm, portanto, do projeto AToLE.

O genoma do organismo estudado, *Angomonas ambiguus*, foi sequenciado em nível rascunho através da plataforma Illumina MiSeq. Foi gerado, a partir de uma biblioteca TruSeq LT, um total de 13.233.416 de leituras (pareadas, 2 x 250). A montagem foi feita utilizando-se o programa Newbler v. 2.9 (Roche 454), com parâmetros *default* (exceto pelo uso de *-large*, necessário devido ao número de repetições no genoma, que torna a montagem muito lenta).

As sequências pertencentes ao endossimbionte foram inicialmente separadas das nucleares por similaridade com os genomas, previamente sequenciados completamente, de outras cinetoplastibactérias (ALVES et al., 2013b). Com isso, foram obtidos 14 *contigs* perfazendo cerca de 816 kbp de sequência genômica (bem próximo aos valores vistos nos outros TPEs, entre 810 kbp e 833 kbp). No entanto,

esse processo ainda pode potencialmente deixar de capturar algumas sequências, únicas ao TPE de *A. ambiguus* e presentes em *contigs* menores.

### 3.2 PREDIÇÃO GÊNICA

Para predição das sequências gênicas, tanto para o genoma do tripanossomatídeo quanto para a  $\beta$ -proteobactéria endossimbionte, foi utilizado Glimmer v.3.02 (DELCHER et al., 2007). O Glimmer é uma coleção de programas para identificação de genes de bactérias, o qual usa um conjunto de genes de treinamento para criação de um modelo de Markov. Foram utilizados os parâmetros padrões, tanto para o genoma procarioto quanto para o eucarioto. Nesta etapa foram preditos os genes para os organismos eucarioto e procarioto. Regiões correspondentes a RNAs ribossômicos e transportadores foram preditos usando os programas RNAmmer v. 1.2 (LAGESEN et al., 2007) e tRNAscan-SE v. 1.4 (LOWE; EDDY, 1997), respectivamente.

Além dos genomas de *A. ambiguus* e seu endossimbionte, foi realizada a predição de sequências codificantes dos genomas de outros cinco tripanossomatídeos e seus respectivos TPEs, uma vez que os dados destes genomas serão necessários para a comparação genômica e análise filogenética. São eles: *Angomonas deanei* e *Candidatus Kinetoplastibacterium crithidii*; *Angomonas desouzai* e *Ca. K. desouzaii*; *Strigomonas culicis* e *Ca. K. blastocrithidii*; *Strigomonas galati* e *Ca. K. galatii*; *Strigomonas oncopelti* e *Ca. K. oncopeltii*.

### 3.3 ANOTAÇÃO

A primeira etapa da anotação das sequências preditas foi realizada por similaridade através do programa BLAST e rpsBLAST v. 2.4.0+ (CAMACHO et al., 2009), contra os bancos de dados NR (*Non-redundant*) e CDD (*Conserved Domains Database*), respectivamente, disponíveis no NCBI (<ftp://ftp.ncbi.nih.gov/blast/db/>).

Para anotação metabólica utilizou-se o ASGARD v.1.6.5 (ALVES; BUCK, 2007), usando termos do KEGG Orthology e Uniprot. A partir destes bancos de dados, o ASGARD lida com as sequências proteicas preditas em etapas anteriores para anotá-las, identificando as vias metabólicas de que participam, além de permitir a comparação de mapas metabólicos entre espécies, identificando vias que espécies podem compartilhar ou não. Também utilizou-se o InterProScan v5.20-59.0, com intuito de buscar motivos proteicos em várias bases de domínios e/ou famílias de proteínas.

Através do Artemis (CARVER et al., 2008), um visualizador de genomas, foi possível realizar a correção de erros de quebra de ORFs, bem como a eliminação de pseudogenes preditos equivocadamente.

### 3.4 GENÔMICA COMPARATIVA

Uma das possíveis formas de comparar os genomas entre espécies é a partir da identificação de genes ortólogos. A identificação dos grupos de genes ortólogos foi realizada através do programa OrthoMCL v.2.0 (LI et al., 2003). Este programa utiliza-se do BlastP para fazer um alinhamento todos-contra-todos, então as sequências que tiveram melhores *hits* recíprocos são identificadas como possíveis pares ortólogos; os endoparálogos são identificados quando sequências dentro do

mesmo genoma são mutualmente mais semelhantes entre si do que quaisquer outras sequências dos outros genomas. Em seguida, as relações dos possíveis ortólogos e parálogos são convertidos em grafos, nos quais os nós representam proteínas e as arestas são as relações entre eles, agrupados baseados na similaridade entre as sequências. É dada uma pontuação para cada uma das relações de sequências, a qual é representada por uma matriz de pontuação simétrica; por último, é aplicado o algoritmo MCL, o qual vai considerar todas as relações no grafo, separando parálogos divergentes, ortólogos distantes que foram agrupados erroneamente, e sequências com diferentes domínios. Uma grande vantagem do OrthoMCL, é que ele pode ser usado com um grande número de espécies.

Após rodar o OrthoMCL, foram compiladas as sequências de genes ortólogos, sendo que cada grupo possuía sempre seis sequências, uma de cada espécie em estudo; esta separação foi realizada através de scripts em Perl e Bash. Cada grupo foi colocado em um arquivo “.fasta”, com o alinhamento de cada um dos arquivos por meio do programa MUSCLE v.3.8.31 (EDGAR, 2004), utilizando parâmetros padrões; foi criado um *script* em *shell* para automatizar a tarefa. A ferramenta OrthoVenn foi usada para criação do diagrama de Venn das seis espécies de hospedeiros e endossimbiontes (<http://www.bioinfogenome.net/OrthoVenn/>).

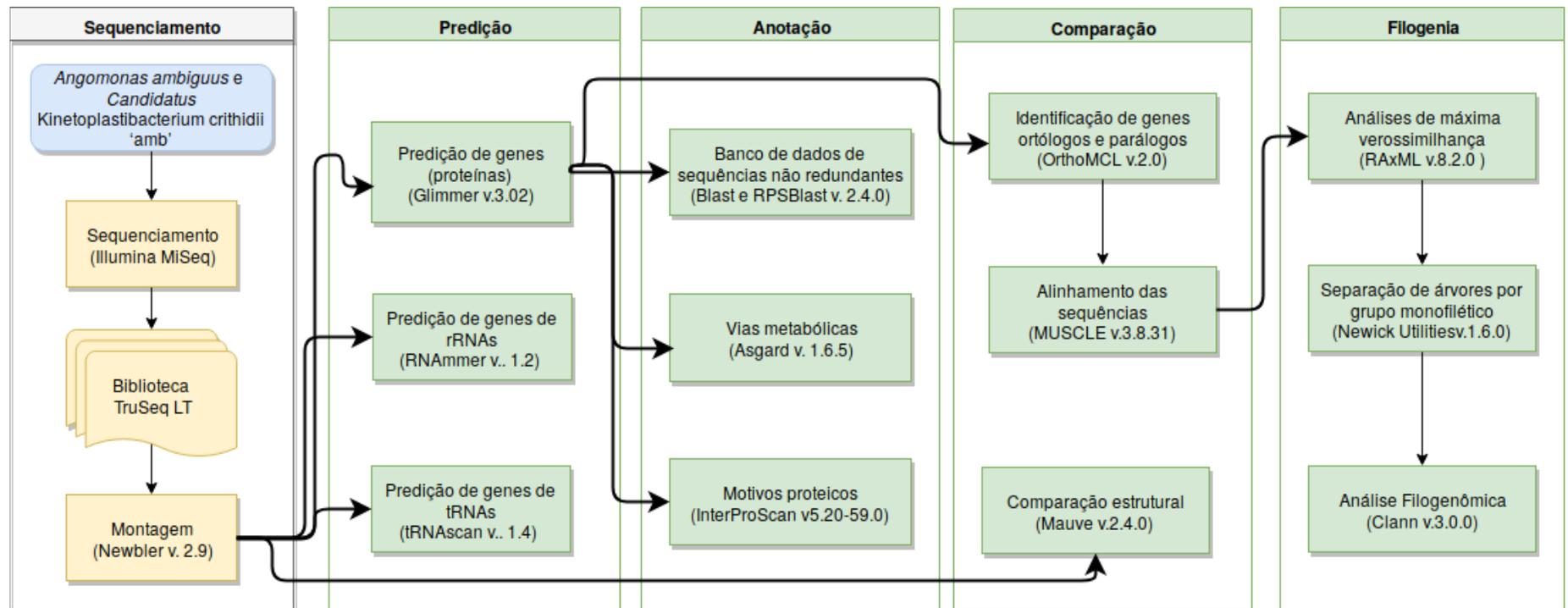
A comparação estrutural entre os genomas foi feita através de análises de rearranjos e sintenia, utilizando o programa Mauve v.2.4.0 (DARLING et al., 2004), tanto para os genomas do núcleo quanto para o do endossimbionte, utilizando a opção *progressive Mauve*, que usa um algoritmo de busca de similaridade mais apurada e sensível, além de que pode ser comparado um número maior de taxa na mesma comparação.

### 3.5 ANÁLISE FILOGENÉTICA

Posteriormente à fase de alinhamento, análises de máxima verossimilhança foram executadas usando o RAxML v.8.2.9 (STAMATAKIS, 2014), utilizando seleção de modelo de substituição (-m) automática (PROTGAMMAAUTO). O suporte de ramos foi estimado pela análise por *bootstrap* (-f a), e o número de pseudoreplicatas foi calculado automaticamente (-# autoMR).

As árvores filogenéticas obtidas foram categorizadas de acordo com os grupos monofiléticos a partir de sequências de *Angomonas ambiguus* e seu TPE, usando como grupo externo *Strigomonas culicis* e *Ca. K. blastocritidii* para a análise do genoma do tripanossomatídeo e do endossimbionte, respectivamente; esta etapa foi feita com o programa Newick Utilities v.1.6.0 (RAGONNET-CRONIN et al., 2013), utilizando as árvores no formato newick. Além disso, usou-se o Clann v.3.0.0 (CREEVEY; MCINERNEY, 2009) para a construção da superárvore, utilizando parâmetros padrões, primeiramente através da análise de *bootstrap*, e então fazendo uma busca heurística para busca da melhor árvore consenso. Um fluxograma com as etapas desenvolvidas neste trabalho é encontrado na Figura 4.

Fig 4. Fluxo de etapas desenvolvidas neste projeto.



\*etapa de sequenciamento e montagem, não foi realizada nessa pesquisa.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 CARACTERIZAÇÃO GERAL DOS GENOMAS

Foram utilizados dois genomas principais, o de *Angomonas ambiguus* e o do seu endossimbionte, *Ca. K. crithidii* 'amb', além dos genomas das cinco espécies relacionadas, juntamente com os seus respectivos TPEs, para a análise comparativa.

Um total de 10.911 genes codificadores de proteínas, 7 sequências de rRNAs e 153 de tRNAs foram preditos para *A. ambiguus*. Dos sete rRNAs, seis são 8S e um 18S. Para *Ca. K. crithidii* 'amb', identificamos 755 genes, 3 rRNAs (16s, 5s e 23s) e 41 tRNAs. Os dados obtidos a partir da montagem e predições são mostrados a seguir, comparativamente às espécies *Angomonas deanei*, *Angomonas desouzai*, *Strigomonas culicis*, *Strigomonas galati* e *Strigomonas oncopelti* (Tabela 1) e seus TPEs (Tabela 2), lembrando que o genoma de *Ca. K. crithidii* 'amb' não teve sua sequência finalizada em um só *contig* e, portanto, alguns dos números são preliminares.

**Tabela 1:** Dados comparativos dos genomas dos SHTs

	A. amb	A. dea	A. des	S. cul	S. gal	S. onc
Tamanho do genoma (Mb)	21,194	23,079	24,250	23,587	27,235	24,96
Média GC (%)	46,10	49,24	49,15	56,72	49,94	54,94
CDS	14,239	17,952	20,905	15,317	14,868	19,022
Tamanho médio CDS (kb)	943,47	882,55	698,67	1,003,39	990,41	671,75
Genes de rRNA	7	20	17	14	17	15
Genes de tRNA	153	97	109	140	149	88

**Tabela 2:** Dados comparativos dos genomas dos TPEs

	C. K. cri 'amb'	C. K. cri	C. K. des	C. K. bla	C. K. gal	C. K. onc
Tamanho do genoma (kb)	803,474	821,930	833,125	820,029	822,140	810,170
Média GC (%)	30,65	30,29	30,17	32,55	32,36	31,23
CDS	754	781	794	776	779	746
Tamanho médio CDS (kb)	948.40	964.31	961.94	957.28	965.85	967.58
Genes de rRNA	9	9	9	9	9	9
Genes de tRNA	41	44	43	43	43	43

Os genomas dos tripanossomatídeos portadores de endossimbiontes (SHTs) estudados são similares entre si, apresentando pouca variação no tamanho do genoma, no conteúdo GC e número de genes. A única divergência que vemos é o número de rRNAs exibido por *Angomonas ambiguus*, que apresenta somente sete genes de rRNAs, enquanto as outras espécies possuem entre 14 e 20 sequências.

O mesmo acontece nos genomas dos endossimbiontes, *Ca. K. crithidii* 'amb' apresenta somente três genes de rRNAs, enquanto todos os outros possuem nove sequências. Outros indicadores, como tamanho do genoma, conteúdo GC, número de genes e número de tRNAs, são similares nas seis espécies de  $\beta$ -proteobactérias, semelhantemente ao apresentado no trabalho de Motta et al. (2013).

Tanto *A. ambiguus* quanto seu endossimbionte apresentaram números menores de genes para RNA ribossômico, quando comparamos com as outras espécies em estudo. Isso se deve provavelmente a diferenças na maneira em que os genomas foram sequenciados e conseqüentemente na montagem dos mesmos. O genoma de *A. ambiguus* foi sequenciado através da plataforma da Illumina (MiSeq). Características desta tecnologia são gerar *reads* curtos, comparados com 454 (Roche), usado para os outros organismos, e ter uma alta cobertura. Isso pode causar um problema na montagem. Quando existem diferentes cópias idênticas do gene e as leituras não conseguem abranger o tamanho total da região repetitiva, no processo de montagem as cópias dos genes formadas podem ser colapsadas em menos *contigs*, mostrando um número menor de sequências do que verdadeiramente existe.

Os genomas dos endossimbiontes são visivelmente menores comparados aos de outras  $\beta$ -proteobactérias de vida livre, como por exemplo *Candidatus Accumulibacter*, que possui um genoma de 4,48 Mb (NCBI); os genomas das espécies de  $\beta$ -proteobactéria em estudo possuem por volta de 800 kb (Fig 3), o que nos revela uma deterioração de conteúdo genômico ao longo do tempo devido à

endossimbiose. Diversos trabalhos revelam a diminuição de conteúdo genômico em espécies endossimbiontes principalmente devido à perda de genes regulatórios, incluindo reguladores de transcrição, genes de reparo de DNA e genes para a síntese componentes de superfície celular (MORAN; MCCUTCHEON; NAKABACHI, 2008), o que certamente ocorreu com os genomas dos endossimbiontes em estudo.

## 4.2 ANÁLISE DO CONTEÚDO GÊNICO

A identificação de genes ortólogos através do OrthoMCL permite realizar uma comparação entre todos os genes ortólogos compartilhados entre os TPEs (além dos hospedeiros), como pode ser resumido na Fig. 5. Nela vemos a presença e ausência de genes codificadores de proteínas ordenados e alinhados a partir do primeiro (*dnaA*) ao último gene.

Constatamos um grande agrupamento de blocos brancos (perda de genes) na mesma região do genoma. *Ca. K. crithidii* ‘amb’ apresenta quatro genes ortólogos presente somente em seu genoma (bloco verde), enquanto *Ca. K. crithidii* tem três, *Ca. K. desouzaii*, dez; *Ca. K. blastocrithidii* tem um, *Ca. K. galatii* e *Ca. K. oncopeltii* nenhum. Enquanto pseudogenes são bastante encontrados nos genomas de TPEs de *Strigonomas* (24), quando comparados aos genomas de TPEs de *Angomonas* (5).

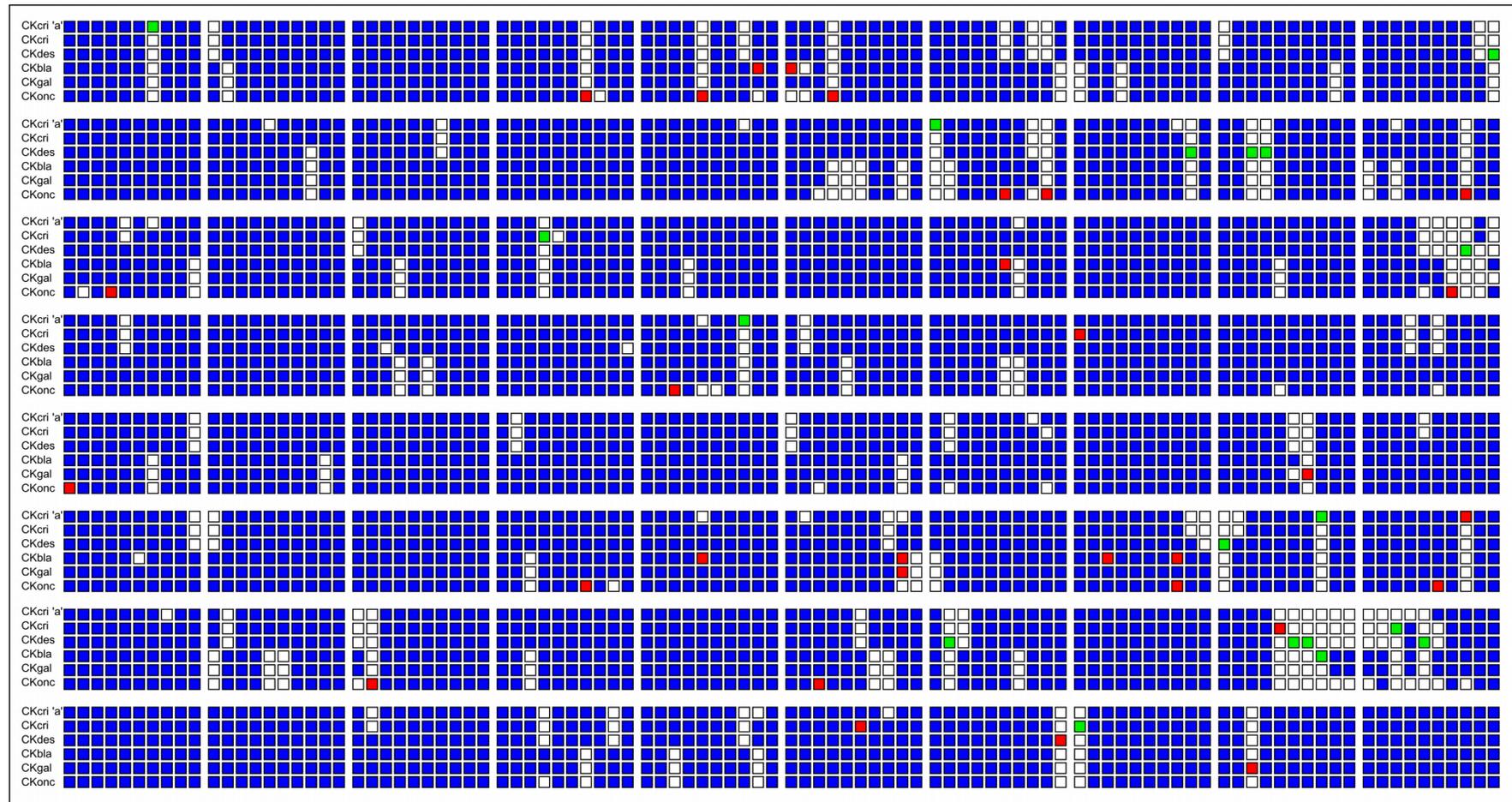
A perda de genes parece ser comum em endossimbiontes e explicada pelo relaxamento da seleção natural (*relaxed selection*). Esta seleção é referida quando a fonte de seleção é enfraquecida ou removida, e ela pode acontecer, por exemplo, quando há uma redução do tamanho populacional (WERTHEIM *et al.*, 2014). Ela aponta padrões marcantes de perdas de genes entre endossimbiontes, como a tendência de perda de genes para vias metabólicas, mas mantêm as funções básicas

de manutenção. Este processo também pode explicar a colaboração de metabólitos que são benéficos ao hospedeiro, devido à permanência de algumas vias específicas de biossíntese (WERNEGREN, 2015). A perda de gene ocorre principalmente pela acumulo de substituições que transformam um gene em pseudogene, posteriormente perdendo a sequência totalmente durante a replicação (ALVES *et al.*, 2013b).

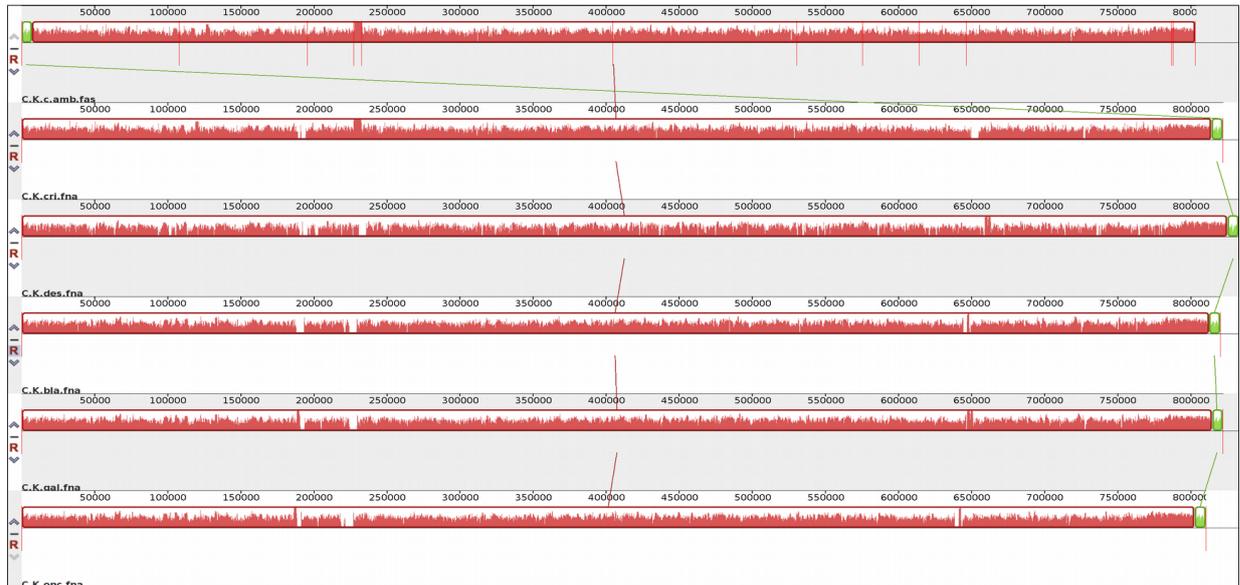
Apesar das pequenas diferenças vistas quando comparamos o conteúdo genético dos genomas dos TPEs, quando fazemos uma equiparação mais bruta, como o alinhamento do genoma inteiro, um contra o outro, percebemos que as diferenças diminuem. Aqui, os genomas das seis espécies de endossimbiontes foram alinhados através do programa Mauve, do qual foi obtido o seguinte resultado (Fig. 6). Como a recombinação pode causar rearranjos no genoma, as regiões ortólogas de um genoma podem ser reordenadas ou invertidas em relação a outro genoma. Por isso realizar o alinhamento do genoma por inteiro de duas ou mais espécies pode levantar questões interessantes sobre rearranjos e inferir relações evolutivas entre eles.

Espécies que são relacionadas evolutivamente, como os TPEs, possuem uma conservada sintenia, com genes vizinhos tendendo a manter suas posições e ordens ao longo do tempo evolutivo. Corroborando dados anteriores, vemos na Fig. 6 que o genoma de *Ca. K. crithidii* 'amb', quando fazemos uma comparação do genoma inteiro com os outros genomas de TPEs, é bem semelhante em termos de organização, mostrando um grande grau de colinearidade com aqueles.

**Fig. 5:** Representação dos grupos ortólogos no genoma das seis espécies de endossimbiontes *Candidatus* Kinetoplastibacterium em ordem de ocorrência após o gene *dnaA*. Caixas na mesma coluna representam genes ortólogos identificados pelo OrthoMCL. Caixas azuis representam ortólogos que estão presentes em mais de um organismo; verdes, ortólogos presentes em único organismo (*singleton*); vermelhos, pseudogenes; brancos, ortólogos ausentes na espécie. Abreviaturas: CKcri ‘a’ denota *Ca. K. crithidii* ‘amb’; CKcri, *Ca. K. crithidii*; CKdes, *Ca. K. desouzaii*; CKbla, *Ca. K. blastocrithidii*; CKgal, *Ca. K. galatii*; e CKonc, *Ca. K. oncopeltii*.



**Fig. 6:** Comparação linear entre os genomas dos seis endossimbiontes na ordem *Ca. K. cri* ‘amb’, *Ca. K. cri*, *Ca. K. des*, *Ca. K. blas*, *Ca. K. gal*, *Ca. K. onc*.



A única observação a ser feita no alinhamento dos genomas, é a pequena região em verde, presente em todas as espécies, em *Ca. K. crithidii* ‘amb’ essa porção aparece no começo, e nas outras espécies no final do genoma, isso se deve ao primeiro *contig* do genoma de *Ca. K. crithidii* ‘amb’ não começar com o *dnaA*.

### 4.3 ANOTAÇÃO

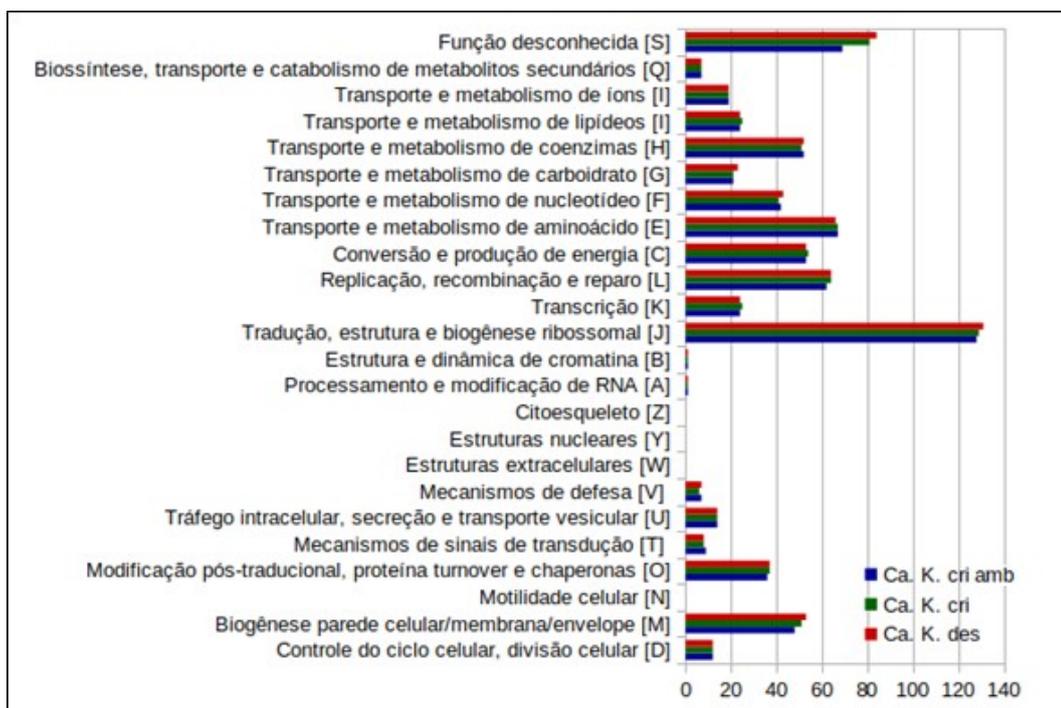
A primeira parte da anotação de todos os genes presentes em *A. ambiguus* e *Ca. K. crithidii* ‘amb’, foi realizada usando o BLAST (blastp, principalmente) em busca de sequências com maior similaridade nos bancos de dados NR e CDD, para posterior publicação de dados no GenBank. A outra parte da anotação foi realizada a partir o EggNOG, vista a seguir, e ASGARD, para identificação de vias metabólicas.

#### - Comparação funcional de proteínas em TPEs

A utilização do EggNOG permite anotar sequências gênicas codificadoras de proteínas, de acordo com 25 categorias funcionais, que proporcionam uma visão mais ampla dos processos desempenhados pelas proteínas. A Fig. 7 resume os

dados de saída do EggNOG 4.5.1 (<http://eggnogdb.embl.de/#/app/home>) de todas as proteínas três TPEs de tripanossomatídeos do gênero *Angomonas*, *Ca. K. crithidii* ‘amb’, *Ca. K. crithidii*, *Ca. K. desouzai*.

**Fig. 7:** Categorização funcional de proteínas nos genomas de três endossimbiontes do gênero *Angomonas*.



Como é esperado encontrar em genomas de bactérias, a grande maioria das proteínas encontradas está na categoria “Tradução, estrutura e biogênese ribossomal”, logo em seguida em famílias de proteínas que não possuem uma função conhecida, depois nas categorias “Transporte e metabolismo de aminoácidos” e “Replicação, recombinação e reparo”.

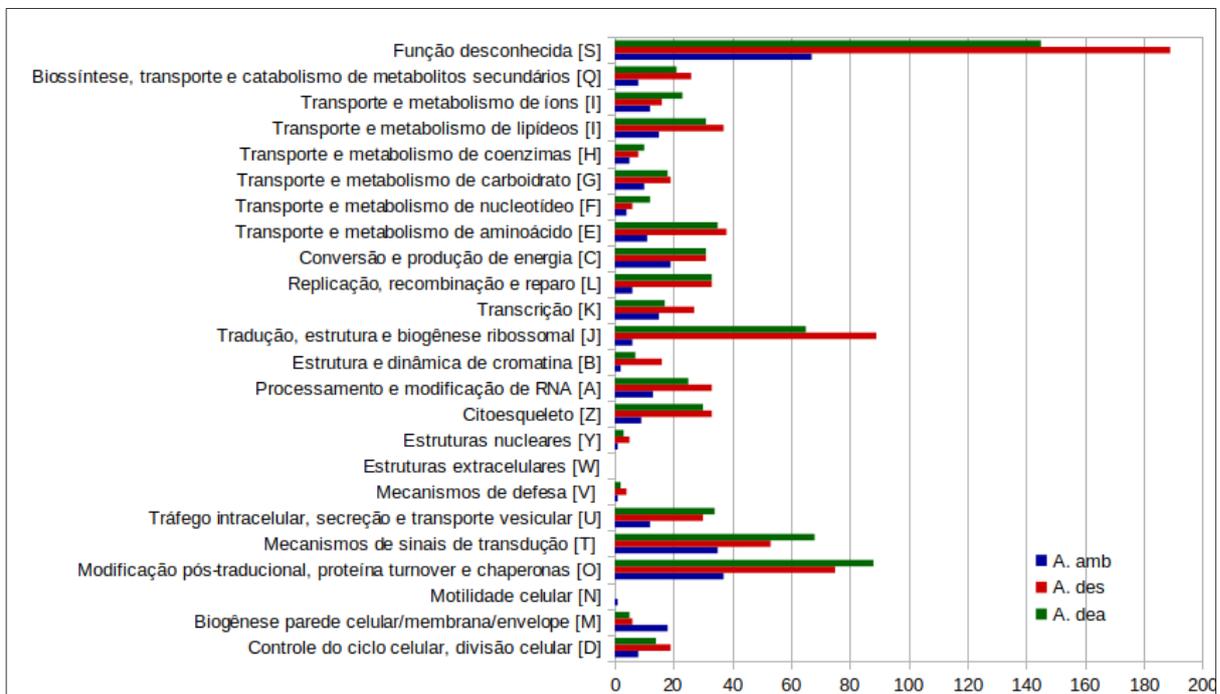
Grande parte dos genes está envolvida na manutenção do próprio genoma e na codificação de genes ribossomais. Apesar da diminuição do conteúdo genômico nesses genomas, vemos uma pressão seletiva na permanência de genes que estão envolvidos na relação simbiótica entre TPE e SHT, vista na permanência de grande

quantidade de proteínas envolvidas no metabolismo de aminoácidos, que como veremos a seguir é uma das principais contribuições da bactéria para o eucarioto.

- *Genes exclusivos de A. ambiguus*

Para os genomas dos SHTs do gênero *Angomonas* (*ambiguus*, *desouzai* e *deanei*) comparamos as categorias proteicas de genes que são únicos para cada espécie (segundo a saída do programa OrthoMCL), visto na Fig. 8.

**Fig. 8:** Genes únicos de cada espécie de SHT do gênero *Angomonas* categorizados pelo produto proteico no eggNOG.



Podemos perceber que *A. desouzai* (cor vermelha) apresenta, quase sempre, a maior quantidade de genes únicos nas categorias apresentadas, excetuando as U, T e O, onde *A. deanei* apresenta o maior número de genes, enquanto *A. ambiguus*, revela uma quantidade bem menor de genes únicos, com exceção da categoria M (Biogênese parede celular/membrana/envelope). Indicar as características de sequências presentes somente em um organismo pode levantar questões biológicas

importantes a respeito do mesmo, e tais dados podem ser úteis para futuras pesquisas, inclusive para comprovar os dados *in silico*.

#### 4.2.1 KINETOPLASTO

Identificamos três *contigs* no genoma de *A. ambiguus* nos quais estão presentes os genes de DNA mitocondrial. As sequências gênicas de maxicírculos são muito bem conservadas na família Trypanosomatidae, por isso é possível identificá-las por similaridade. Sendo assim foram encontrados 20 genes pertencentes ao maxicírculo em *A. ambiguus*, vistos na Tabela 3, abaixo.

**Tabela 3.** Genes encontrados no maxicírculo de *A. ambiguus* a partir da ferramenta *blast* contra banco de dados NR e espécies identificadas.

Gene	Espécie
ATPase subunit 6	<i>Angomonas deanei</i>
cytochrome b	<i>Angomonas deanei</i>
cytochrome c oxidase subunit I	<i>Angomonas deanei</i>
cytochrome c oxidase subunit II	<i>Angomonas deanei</i>
cytochrome c oxidase subunit III	<i>Angomonas deanei</i>
hypothetical protein	<i>Angomonas deanei</i>
NADH dehydrogenase subunit 1	<i>Angomonas deanei</i>
NADH dehydrogenase subunit 2	<i>Angomonas deanei</i>
NADH dehydrogenase subunit 4	<i>Angomonas deanei</i>
NADH dehydrogenase subunit 5	<i>Angomonas deanei</i>
NADH dehydrogenase subunit 7	<i>Angomonas deanei</i>
C-rich region 4	<i>Trypanosoma vivax</i>
cytochrome oxidase subunit 1	<i>Trypanosoma vivax</i>
cytochrome oxidase subunit II	<i>Trypanosoma vivax</i>
cytochrome oxidase subunit III partial	<i>Trypanosoma vivax</i>
NADH dehydrogenase subunit 3 partial	<i>Trypanosoma vivax</i>
NADH dehydrogenase subunit 8	<i>Trypanosoma vivax</i>
NADH dehydrogenase subunit 9	<i>Trypanosoma vivax</i>
NADH_dehydrogenase subunit 2	<i>Trypanosoma vivax</i>
ribosomal protein S12	<i>Trypanosoma vivax</i>

Percebemos que, dos 20 genes encontrados do maxicírculo de *A. ambiguus*, 11 são mais relacionados a *A. deanei*, mas 9 encontrados foram mais similares aos

genes de *T. vivax*. Como dito anteriormente, as sequências do kinetoplasto são bem conservadas, por isso deram *match* em *T. vivax*, espécie distante filogeneticamente.

#### 4.2.2 VIAS METABÓLICAS

A relação simbiótica mutualística entre os tripanossomatídeos e seus respectivos endossimbiontes traz vantagens metabólicas particulares para os hospedeiros. Um exemplo são as vias metabólicas responsáveis pela síntese de alguns aminoácidos que somente os SHT conseguem fazer. Assim como visto em outros tripanossomatídeos dos gêneros *Angomonas* e *Strigomonas*, *A. ambiguus* e *Ca. K. crithidii* 'amb' também compartilham vias que completam o metabolismo de alguns aminoácidos, como arginina, prolina, valina, leucina, isoleucina, histidina, lisina (descritos abaixo), metionina, cisteína, treonina, fenilalanina, tirosina, triptofano.

##### - *Arginina e Prolina*

O glutamato é o precursor da prolina e da arginina (além da glutamina), neste mapa vemos que o endossimbionte fornece os compostos N-acetil-glutamato, N-acetil-glutamina-P, N-acetil-glutamato semialdeído, N-acetil-ornitina (que o *A. ambiguus* também possui), passando pela ornitina, para a conversão do glutamato em prolina.



Tripanossomatídeos que não são SHT, ou tripanossomatídeos regulares (RT), como o *Trypanossoma cruzi*, *T. brucei* e *Leishmania major*, necessitam importar alguns compostos intermediários da via da prolina para poder sintetizá-la, através de um complexo sistema de proteínas transmembranares (SILBER et al., 2002; BRINGAUD; BARRETT; ZILBERSTEIN, 2012), codificados por 17~21 genes, dependendo da espécie (JACKSON, 2007). Em *A. ambiguus* encontramos somente três desses genes, o que sugere que os genes responsáveis pelo transporte de aminoácidos nesta espécie tenham se perdido pela não mais necessidade de transporte.

A biossíntese de arginina utiliza algumas reações do ciclo da ureia. Neste caso, há ativação e redução do grupo carboxila do glutamato, passando posteriormente à ornitina; convertida, por uma enzima fornecida pelo endossimbionte (2.1.3.3), para citrulina; então transformada para argininossuccinato e por fim, à arginina. Algumas das enzimas desta via apresentadas pelo SHT (Fig.9 - azul), como aquelas que completam o ciclo da ureia, podem ter sido transferidas lateralmente de outras bactérias, como aquelas do filo Firmicutes, durante a evolução da família Trypanossomatidae (ALVES et al., 2013a). No entanto, sem os genes do endossimbionte, *A. ambiguus* não completaria o ciclo de glutamato para arginina, pelo ciclo da ureia.







essencial para o crescimento de RT; em meio de cultura esses tripanossomatídeos não crescem sem a adição de lisina, ao contrário de SHTs, já que os endossimbiontes, incluindo *Ca. K. crithidii* 'amb', fornecem todos os nove genes que codificam enzimas para a síntese de lisina a partir do aspartato.

A última etapa para a síntese de lisina, onde meso-2,6-diaminopelato é convertido em lisina, é feita pela enzima diaminopimelato descarboxilase, codificada por genes apresentados tanto pelo hospedeiro quanto pelo endossimbionte, sugerindo que o SHT também pode controlar a síntese desse aminoácido.

#### 4.2.3 IDENTIFICAÇÃO DE GENES ORTÓLOGOS PARA ANÁLISE FILOGENÉTICA

Foram encontrados 12.829 grupos de genes ortólogos em que pelo menos duas das seis espécies de tripanossomatídeos compartilham genes ortólogos, ou seja, não estão contabilizados os genes que não são compartilhados; 5526 grupos apresentam um ou mais genes compartilhados entre todas as seis espécies de SHTs, e desses somente 4428 grupos (34,51%) foram aqueles que possuíam sequências de cópia única entre as espécies.



Como visão geral do genoma das proteobactérias endossimbiontes, percebe-se que os genomas são bem relacionados, visto que são compartilhados, dos 771 grupos encontrados com o programa OrthoMCL, 641, ou seja, 83,13% dos grupos ortólogos são compartilhados. Apesar do OrthoMCL identificar 43 genes que somente a espécie *Ca. K. crithidii* ‘amb’ possui (singletons), com refinação dos resultados, identificamos 3 singletons desse endossimbionte.

Utilizando somente os grupos ortólogos em que cada uma das seis espécies, tanto de SHTs quanto de TPEs, apresentava um único gene, foram realizados o alinhamento e a construção das árvores, produzindo, portanto, uma árvore para cada grupo de ortólogo. O programa Newick utilities foi utilizado para separar as árvores de cada clado monofilético de *A. ambiguus* e *Candidatus Kinetoplastibacterium crithidii* ‘amb’, descritas nas tabelas 4 e 5, respectivamente.

**Tabela 4:** N° de árvores em que cada espécie aparece em um mesmo clado com *Angomonas ambiguus*.

Espécie	N° de árvores
<i>Angomonas desouzai</i>	4.126
<i>Angomonas deanei</i>	135
<i>Strigomonas galati</i>	0
<i>Strigomonas culicis</i>	0
<i>Strigomonas oncopelti</i>	0

**Tabela 5:** N° de árvores em que cada espécie aparece em um mesmo clado com *Candidatus Kinetoplastibacterium crithidii* ‘amb’.

Espécie	N° de árvores
<i>Candidatus Kinetoplastibacterium desouzaii</i>	4
<i>Candidatus Kinetoplastibacterium crithidii</i>	633
<i>Candidatus Kinetoplastibacterium galatii</i>	0
<i>Candidatus Kinetoplastibacterium blastocrithidii</i>	0
<i>Candidatus Kinetoplastibacterium oncolpeltii</i>	0

Na tabela 4 pode ser visto que do total de 4.428 árvores construídas, 4.126 (93,17%) apresentavam *A. ambiguus* e *A. desouzai* agrupadas em um mesmo clado, enquanto 135 árvores (3,04%) apresentavam *A. ambiguus* agrupada com *A. deanei*. Já para as espécies do gênero *Strigomonas*, nenhuma das árvores encontradas apresentava algum agrupamento direto com *A. ambiguus*.

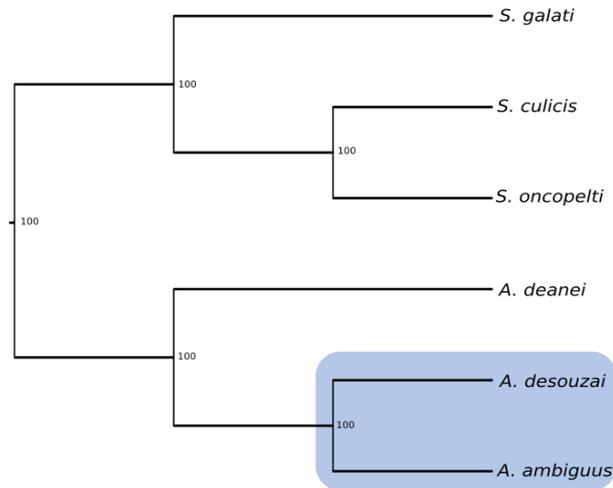
Os resultados da análise realizada para os endossimbiontes são observados na tabela 5. *Ca. K. crithidii* ‘amb’ apareceu em um mesmo clado junto a *Ca. K. desouzaii* em 4 (0,62%) das 641 árvores do total, enquanto que 633 (98,75%) destas mostraram *Ca. K. crithidii* em um clado unicamente com *Ca. K. crithidii* ‘amb’. Nesta análise é possível ver a estreita relação dos endossimbiontes *Ca. K. crithidii* ‘amb’ e *Ca. K. crithidii* filogeneticamente, criando uma discordância entre os genomas presentes em *Angomonas ambiguus*: enquanto o núcleo é aparentado a *A. desouzai*, o genoma do endossimbionte está mais relacionado ao de *Ca. K. crithidii*, endossimbionte de *A. deanei*, corroborando os dados de Teixeira e col. (2011)

A análise filogenética realizada por Teixeira e col. (2011), separou os cladogramas dos endossimbiontes de *Strigomonas* e *Angomonas* em dois pelas diferenças nas sequências no gene SSU rRNA (~13%) e gGAPDH (~12%), além de ser bem suportado para a separação dos subclados internos.

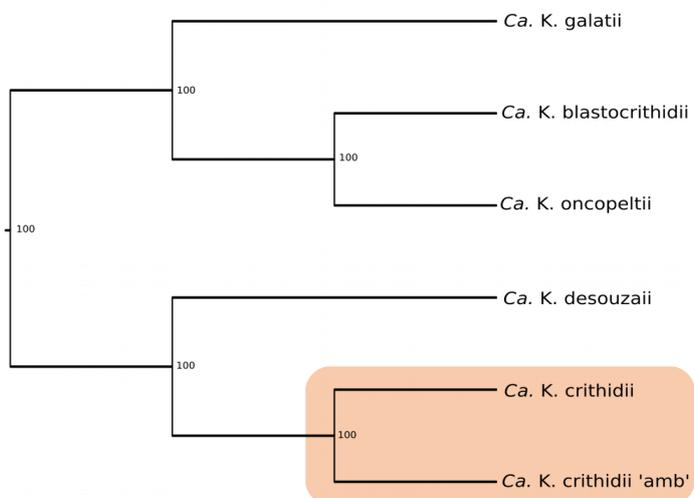
#### 4.4 FILOGENÔMICA

A análise filogenômica baseada em superárvore confirmou os dados mostrados anteriormente (Fig. 15 e 16).

**Fig. 15.** Árvore filogenética construída a partir das 4.428 árvores resultantes da análise de comparação de ortólogos dos tripanossomatídeos por meio da abordagem superárvore. Valores de *bootstrap* presentes em cada nó.



**Fig 16** Árvore filogenética construída a partir das 641 árvores resultantes da análise de comparação de ortólogos dos endossimbiontes por meio da abordagem superárvore. Valores de *bootstrap* presentes em cada nó.



A Fig. 15 representa os tripanossomatídeos que albergam bactérias endossimbiontes. O gênero *Strigomonas* aparece em um grupo monofilético, enquanto o gênero *Angomonas* está em outro ramo do cladograma. Dentro deste clado, *A. ambiguus* está diretamente relacionada com *A. desouzai* formando outro grupo monofilético. Já na Fig. 16, representando as relações dos endossimbiontes desses tripanossomatídeos, vemos *Ca. K. crithidii* ‘amb’ em um mesmo clado junto a *Ca. K. crithidii*, o TPE de *A. deanei*.

Tratando-se da relação filogenética entre as espécies, a análise dos seis tripanossomatídeos entra em acordo com Borghesan et al. (2013), separando *Angomonas* e *Strigomonas* em dois ramos distintos, e com *A. ambiguus* mais relacionada com *A. desouzai*. Naquele trabalho, foram utilizados como marcadores moleculares os genes para SSU rRNA e gGAPDH para determinação da divisão dos tripanossomatídeos portadores de endossimbiontes em dois gêneros, enquanto que na presente pesquisa obtivemos os mesmos resultados através de uma análise filogenômica.

É bem relatado que em espécies de *Leishmania* existem casos de recombinação entre linhagens distintas (TIBAYRENC; AYALA, 1999), e até espécies diferentes (SINGH; CHIKARA; SUNDAR, 2013). Kelly e colaboradores (1991), a partir de um estudo de cariotipagem molecular de duas cepas de *Leishmania* isoladas de insetos na Arábia Saudita, isolaram uma cepa distinta das outras espécies, com características de *L. major* e *L. arabica*; outro trabalho identificou isolados de híbridos de *Leishmania* em humanos no Equador, possível cruzamento entre *L. brazilienses* e *L. guyanensis* (BAÑULS et al., 1997).

Como pode ser visto, a grande maioria dos relatos são envolvendo TR, já que grande parte dos centros de pesquisa envolvendo estudos de tripanossomatídeos está voltada para essa pequena amostra, aqueles que não possuem endossimbionte

mas possuem algum interesse de saúde pública e econômico. Entretanto, é bem possível que a mesma situação de hibridização tenha acontecido com as espécies *A. deanei* e *A. desouzai*, visto que eles parasitam normalmente as mesmas espécies de dípteros – o gênero *Angomonas* parasita dípteros da família Calliphoridae, sendo os gêneros *Chrysomya* e *Lucilia* os principais hospedeiros desses tripanossomatídeos (BORGHESAN, 2013). Além disso, como visto neste trabalho, possuem genomas bem similares, portanto com o processo hibridização entre as duas espécies poderia ter surgido uma nova espécie, *A. ambiguus*, permanecendo o endossimbionte de uma espécie e o genoma de outra.

## 5 CONCLUSÃO

- Os genomas de *Angomonas ambiguus* e endossimbionte, *Ca. K. crithidii* ‘amb’ foram analisados e descritos. O genoma do endossimbionte apresenta características comuns aos descritos de outras espécies: perda de conteúdo GC, diminuição no tamanho do genoma pela perda de genes;
- *Ca. K. crithidii* ‘amb’ fornece genes que completam vias de aminoácidos essenciais, ou não, no metabolismo de *A. ambiguus*, mostrando como é importante essa relação simbiótica, tanto para o endossimbionte quanto para o hospedeiro;
- 
- Confirmado por inferência filogenética multigênica, que há uma divergência entre a evolução do genoma do tripanossomatídeo estudado, *Angomonas ambiguus*, e do seu endossimbionte. *A. ambiguus* está mais próxima da espécie *A. desouzai*, enquanto o endossimbionte da primeira está mais próximo do TPE *Ca. Kinetoplastibacterium crithidii*, o qual é endossimbionte de *A. deanei*. Uma explicação para esse caso é a possível hibridação de espécies diferentes, no caso *A. deanei* e *A. desouzai*, que é comum na família Trypanosomatidae.

## REFERÊNCIAS

ABBY, Sophie; DAUBIN, Vincent. Comparative genomics and the evolution of prokaryotes. *Trends in Microbiology*, v. 15, n. 3, p. 135–141, mar. 2007.

ALFIERI, S. C.; CAMARGO, E. P. Trypanosomatidae: isoleucine requirement and threonine deaminase in species with and without endosymbionts. *Experimental Parasitology*, v. 53, n. 3, p. 371–380, jun. 1982.

ALVES, J. M. P. *et al.* Identification and Phylogenetic Analysis of Heme Synthesis Genes in Trypanosomatids and Their Bacterial Endosymbionts. *PLoS ONE*, v. 6, n. 8, p. e23518, 2011.

ALVES, J. M. *et al.* Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evolutionary Biology*, v. 13, p. 190, 9 set. 2013-a.

ALVES, J. M. P. *et al.* Genome Evolution and Phylogenomic Analysis of Candidatus Kinetoplastibacterium, the Betaproteobacterial Endosymbionts of Strigomonas and Angomonas. *Genome Biology and Evolution*, v. 5, n. 2, p. 338–350, 2013-b.

ALVES, João M P; BUCK, Gregory A. Automated system for gene annotation and metabolic pathway reconstruction using general sequence databases. *Chemistry & Biodiversity*, v. 4, n. 11, p. 2593–2602, nov. 2007.

BAÑULS, A. L. *et al.* Evidence for hybridization by multilocus enzyme electrophoresis and random amplified polymorphic DNA between *Leishmania braziliensis* and *Leishmania panamensis/guyanensis* in Ecuador. *The Journal of Eukaryotic Microbiology*, v. 44, n. 5, p. 408–411, out. 1997.

BESEMER, John; BORODOVSKY, Mark. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*, v. 33, n. Web Server issue, p. W451–W454, 1 jul. 2005.

BORGHESAN, Tarcilla Corrente. *Diversidade e filogenia de tripanossomatídeos parasitas de dípteros*. 2013. text – Universidade de São Paulo, 2013. Disponível em:

<<http://www.teses.usp.br/teses/disponiveis/42/42135/tde-03062014-105719/>>.

Acesso em: 4 fev. 2017.

BREGLIA, S. A.; YUBUKI, N.; LEANDER, B. S. Ultrastructure and Molecular Phylogenetic Position of *Heteronema scaphurum*: A Eukaryovorous Euglenid with a Cytoproct. *Journal of Eukaryotic Microbiology*, v. 60, n. 2, p. 107–120, 1 mar. 2013.

BRINGAUD, Frederic; BARRETT, Michael P.; ZILBERSTEIN, Dan. Multiple roles of proline transport and metabolism in trypanosomatids. *Frontiers in Bioscience (Landmark Edition)*, v. 17, p. 349–374, 1 jan. 2012.

BROSCH, R. *et al.* The evolution of mycobacterial pathogenicity: clues from comparative genomics. *Trends in Microbiology*, v. 9, n. 9, p. 452–458, set. 2001.

CAMACHO, Christiam *et al.* BLAST+: architecture and applications. *BMC Bioinformatics*, v. 10, n. 1, p. 421, 15 dez. 2009.

CARVER, Tim *et al.* Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics*, v. 24, n. 23, p. 2672–2676, 1 dez. 2008.

CAVALIER-SMITH, T. Eukaryote kingdoms: seven or nine? *Bio Systems*, v. 14, n. 3–4, p. 461–481, 1981.

CAVALIER-SMITH, T. Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biology Letters*, v. 6, n. 3, p. 342–345, 23 jun. 2010.

CHANG, K. P. Reduced growth of *Blastocrithidia culicis* and *Crithidia oncopelti* freed of intracellular symbiotes by chloramphenicol. *The Journal of Protozoology*, v. 22, n. 2, p. 271–276, maio 1975.

CREEVEY, Christopher J.; MCINERNEY, James O. Trees from trees: construction of phylogenetic supertrees using clann. *Methods in Molecular Biology (Clifton, N.J.)*, v. 537, p. 139–161, 2009.

BAKER A. C. Flexibility and Specificity in Coral-Algal Symbiosis: Diversity, Ecology, and Biogeography of Symbiodinium. *Annual Review of Ecology, Evolution, and Systematics*, v. 34, p. 661-689, 2003.

BAZINET, C. Endosymbiotic origins of sex. *BioEssays*, v. 26, n. 5, p. 558–566, 1 maio 2004.

BESEMER, John; BORODOVSKY, Mark. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. *Nucleic Acids Research*, v. 33, n. Web Server issue, p. W451–W454, 1 jul. 2005.

BRANDÃO, A. A. *et al.* The heterogeneity of choanomastigote-shaped trypanosomatids as analyzed by their kDNA minicircle size: taxonomic implications. *Parasitology Research*, v. 86, n. 10, p. 809–812, set. 2000.

D H BOUCHER; S JAMES; KEELER, AND K. H. The Ecology of Mutualism. *Annual Review of Ecology and Systematics*, v. 13, n. 1, p. 315–347, 1982.

DAUBIN, Vincent; GOUY, Manolo; PERRIÈRE, Guy. A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history. *Genome Research*, v. 12, n. 7, p. 1080–1090, jul. 2002.

D'AVILA-LEVY, C. M. *et al.* Bodo sp., a Free-Living Flagellate, Expresses Divergent Proteolytic Activities from the Closely Related Parasitic Trypanosomatids. *Journal of Eukaryotic Microbiology*, v. 56, n. 5, p. 454–458, 1 set. 2009.

D'AVILA-LEVY, C. M. *et al.* Influence of the endosymbiont of *Blastocritidia culicis* and *Crithidia deanei* on the glycoconjugate expression and on *Aedes aegypti* interaction. *FEMS Microbiology Letters*, v. 252, n. 2, p. 279–286, 15 nov. 2005.

DARLING, Aaron C. E. *et al.* Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research*, v. 14, n. 7, p. 1394–1403, jul. 2004.

DE SOUZA, W.; MOTTA, M. C. M. Endosymbiosis in protozoa of the Trypanosomatidae family. *FEMS Microbiology Letters*, v. 173, n. 1, p. 1–8, 1 abr. 1999.

DE SOUZA, A. P. *et al.* Trypanosoma cruzi: host selenium deficiency leads to higher mortality but similar parasitemia in mice. *Experimental Parasitology*, v. 101, n. 4, p. 193–199, ago. 2002.

DELCHER, Arthur L. *et al.* Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*, v. 23, n. 6, p. 673–679, 15 mar. 2007.

DU, Y.; MASLOV, D. A.; CHANG, K. P. Monophyletic origin of beta-division proteobacterial endosymbionts and their coevolution with insect trypanosomatid protozoa Blastocrithidia culicis and Crithidia spp. *Proceedings of the National Academy of Sciences of the United States of America*, v. 91, n. 18, p. 8437–8441, 30 ago. 1994.

DU, Y.; MCLAUGHLIN, G.; CHANG, K. P. 16S ribosomal DNA sequence identities of beta-proteobacterial endosymbionts in three Crithidia species. *Journal of Bacteriology*, v. 176, n. 10, p. 3081–3084, maio 1994.

EDGAR, Robert C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC bioinformatics*, v. 5, p. 113, 19 ago. 2004.

FARIA E SILVA, P. M. *et al.* Herpetomonas roitmani (Fiorini *et al.*, 1989) n. comb.: a trypanosomatid with a bacterium-like endosymbiont in the cytoplasm. *The Journal of Protozoology*, v. 38, n. 5, p. 489–494, out. 1991.

FITCH, Walter M. Distinguishing Homologous from Analogous Proteins. *Systematic Biology*, v. 19, n. 2, p. 99–113, 1 jun. 1970.

FITCH, W. M. Homology a personal view on some of the problems. *Trends in genetics: TIG*, v. 16, n. 5, p. 227–231, maio 2000.

FIORE, C. L. *et al.* Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends in Microbiology*, v. 18, n. 10, p. 455–463, out. 2010.

FIORINI, J. E. *et al.* Three new species of trypanosomatidae isolated in Alfenas, Minas Gerais, Brazil. *Memórias do Instituto Oswaldo Cruz*, v. 84, n. 1, p. 69–74, mar. 1989.

FRASER, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature*, v. 408, n. 6810, p. 325, nov. 2000.

GALINARI, S.; CAMARGO, E. P. Trypanosomatid protozoa: Survey of acetylornithinase and ornithine acetyltransferase. *Experimental Parasitology*, v. 46, n. 2, p. 277–282, dez. 1978.

GALPERIN, M. Y.; KOONIN, E. V. Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operon disruption. *In Silico Biology*, v. 1, n. 1, p. 55–67, 1998.

GALPERIN, M. Y.; KOONIN, E. V. Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica*, v. 106, n. 1–2, p. 159–170, 1999.

GEE, Henry. Evolution: ending incongruence. *Nature*, v. 425, n. 6960, p. 782, 23 out. 2003.

GIBBS, S. P. The chloroplasts of some algal groups may have evolved from endosymbiotic eukaryotic algae. *Annals of the New York Academy of Sciences*, v. 361, p. 193–208, 1981.

GOEL, Neelam; SINGH, Shailendra; ASERI, Trilok Chand. *A Review of Soft Computing Techniques for Gene Prediction*. Research article. Disponível em: <<https://www.hindawi.com/journals/isrn/2013/191206/>>. Acesso em: 10 out. 2017.

GREIF, G. *et al.* Kinetoplast adaptations in American strains from *Trypanosoma vivax*. *Mutation Research*, v. 773, p. 69–82, mar. 2015.

HEATHER, James M.; CHAIN, Benjamin. The sequence of sequencers: The history of sequencing DNA. *Genomics*, v. 107, n. 1, p. 1–8, 1 jan. 2016.

HOLLAR, L.; MASLOV, D. A. A phylogenetic view on the genus *Phytomonas*. *Molecular and Biochemical Parasitology*, v. 89, n. 2, p. 295–299, nov. 1997.

HONEGGER, R. Simon Schwendener (1829-1919) and the Dual Hypothesis of Lichens. *The Bryologist*, v. 103, n. 2, p. 307–313, 2000.

HUDSON, Andre' O. *et al.* Biosynthesis of lysine in plants: evidence for a variant of the known bacterial pathways. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1721, n. 1, p. 27–36, 18 jan. 2005.

HUGHES, A. L.; PIONTKIVSKA, H. Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) Based on 18S rRNA: Evidence for Paraphyly of Trypanosoma and Six Other Genera. *Molecular Biology and Evolution*, v. 20, n. 4, p. 644–652, 4 jan. 2003.

HUYNEN, M. A. *et al.* Combining data from genomes, Y2H and 3D structure indicates that BolA is a reductase interacting with a glutaredoxin. *FEBS letters*, v. 579, n. 3, p. 591–596, 31 jan. 2005.

JACKSON, Andrew P. Origins of amino acid transporter loci in trypanosomatid parasites. *BMC Evolutionary Biology*, v. 7, p. 26, 23 fev. 2007.

KELLY, J. M. *et al.* Evidence of genetic recombination in Leishmania. *Molecular and Biochemical Parasitology*, v. 46, n. 2, p. 253–263, jun. 1991.

KIVIC, P. A.; WALNE, P. L. An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. *Origins of life*, v. 13, n. 3–4, p. 269–288, 1 mar. 1984.

KONDRASHOV, A. S. Comparative genomics and evolutionary biology. *Current Opinion in Genetics & Development*, v. 9, n. 6, p. 624–629, dez. 1999.

KOSTYGOV, Alexei Y. *et al.* Genome of Ca. Pandoraea novymonadis, an Endosymbiotic Bacterium of the Trypanosomatid Novymonas esmeraldas. *Frontiers in Microbiology*, v. 8, 2017. Disponível em: <<https://www.frontiersin.org/articles/10.3389/fmicb.2017.01940/full>>. Acesso em: 16 out. 2017.

KUTSCHERA, U.; NIKLAS, K. J. Endosymbiosis, cell evolution, and speciation. *Theory in Biosciences*, v. 124, n. 1, p. 1–24, 1 ago. 2005.

KUWAHARA, H. *et al.* Reductive genome evolution in chemoautotrophic intracellular symbionts of deep-sea Calyptogena clams. *Extremophiles*, v. 12, n. 3, p. 365–374, 1 maio 2008.

JENSEN, R. E.; ENGLUND, P. T. Network news: the replication of kinetoplast DNA. *Annual Review of Microbiology*, v. 66, p. 473–491, 2012.

LAGESEN, Karin *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, v. 35, n. 9, p. 3100–3108, 5 jan. 2007.

LEANDER, B. S. Did trypanosomatid parasites have photosynthetic ancestors? *Trends in Microbiology*, v. 12, n. 6, p. 251–258, 1 jun. 2004.

LEANDER, B. S.; FARMER, M. A. Comparative morphology of the euglenid pellicle. I. Patterns of strips and pores. *The Journal of Eukaryotic Microbiology*, v. 47, n. 5, p. 469–479, out. 2000.

LI, Li; STOECKERT, Christian J, Jr; ROOS, David S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, v. 13, n. 9, p. 2178–2189, set. 2003.

LOWE, Todd M; EDDY, Sean R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Research*, v. 25, n. 5, p. 0955-964, 3 jan. 1997.

MADINA, B. R. *et al.* Native mitochondrial RNA-binding complexes in kinetoplastid RNA editing differ in guide RNA composition. *RNA*, v. 20, n. 7, p. 1142–1152, jul. 2014.

MARANDE, W.; LUKES, J.; BURGER, G. Unique mitochondrial genome structure in diplomids, the sister group of kinetoplastids. *Eukaryotic Cell*, v. 4, n. 6, p. 1137–1146, jun. 2005.

MASLOV, Dmitri A. *et al.* Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends in Parasitology*, v. 29, n. 1, p. 43–52, 1 jan. 2013.

MAXAM, A. M.; GILBERT, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, v. 74, n. 2, p. 560–564, fev. 1977.

MERESCHKOWSKY, C. *Über Natur und Ursprung der Chromatophoren im Pflanzenreiche*. S.l.: [s.n.], 1905.

MORAN, N. A. Symbiosis. *Current Biology*, v. 16, n. 20, p. R866–R871, 24 out. 2006.

MORAN, N. A.; MCCUTCHEON, J. P.; NAKABACHI, A. Genomics and Evolution of Heritable Bacterial Symbionts. *Annual Review of Genetics*, v. 42, n. 1, p. 165–190, 2008.

MOTTA, M. C. M. *et al.* Morphological and biochemical characterization of the trypanosomatids *Crithidia desouzai* and *Herpetomonas anglusteri*. *Canadian Journal of Zoology*, v. 69, n. 3, p. 571–577, mar. 1991.

MOTTA, M. C. M. Endosymbiosis in Trypanosomatids as a Model to Study Cell Evolution. *The Open Parasitology Journal*, v. 4, p. 139-147, 2010.

MOTTA, M. C. M. *et al.* Predicting the Proteins of *Angomonas deanei*, *Strigomonas culicis* and Their Respective Endosymbionts Reveals New Aspects of the Trypanosomatidae Family. *PLoS ONE*, v. 8, n. 4, 3 abr. 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3616161/>>. Acesso em: 20 jul. 2015.

MOTTA, M. C. M. *et al.* The Bacterium Endosymbiont of *Crithidia deanei* Undergoes Coordinated Division with the Host Cell Nucleus. *PLoS ONE*, v. 5, n. 8, p. e12415, 26 ago. 2010.

MUNDIM, M. H. *et al.* Simple nutrition of *Crithidia deanei*, a reduviid trypanosomatid with an endosymbiont. *The Journal of Protozoology*, v. 21, n. 4, p. 518–521, out. 1974.

MUNDIM, M. H.; ROITMAN, I. Extra Nutritional Requirements of Artificially Aposymbiotic *Crithidia deanei*. *Journal of Eukaryotic Microbiology*, v. 24, n. 2, p. 329–331, 1 maio 1977.

MUNDIM, M. H. *et al.* Simple nutrition of *Crithidia deanei*, a reduviid trypanosomatid with an endosymbiont. *The Journal of Protozoology*, v. 21, n. 4, p. 518–521, out. 1974.

MUSCATINE, L.; MCNEIL, P. L. Endosymbiosis in Hydra and the Evolution of Internal Defense Systems. *American Zoologist*, v. 29, n. 2, p. 371–386, 1989.

NEWTON, I. L. G. *et al.* The Calyptogena magna Chemoautotrophic Symbiont Genome. *Science*, v. 315, n. 5814, p. 998–1000, 16 fev. 2007.

NIERMAN, W.; EISEN, J. A.; FRASER, C. M. Microbial genome sequencing 2000: new insights into physiology, evolution and expression analysis. *Research in Microbiology*, v. 151, n. 2, p. 79–84, mar. 2000.

NOVY, F. G.; MACNEAL, W. J.; TORREY, H. N. The Trypanosomes of Mosquitoes and Other Insects. *The Journal of Infectious Diseases*, v. 4, n. 2, p. 223–276, 10 abr. 1907.

RAGONNET-CRONIN, Manon *et al.* Automated analysis of phylogenetic clusters. *BMC bioinformatics*, v. 14, p. 317, 6 nov. 2013.

REUTER, Jason A.; SPACEK, Damek; SNYDER, Michael P. High-Throughput Sequencing Technologies. *Molecular cell*, v. 58, n. 4, p. 586–597, 21 maio 2015.

ROITMANA, I.; CAMARGO, E. P. Endosymbionts of trypanosomatidae. *Parasitology Today*, v. 1, n. 5, p. 143–144, nov. 1985.

SCHIMPER, A. T. W. Untersuchungen über die chlorophyllkörner und die ihnen homologen gebilde. *Jahrb. Wiss. Botan*, v. 41, n. 105, p. 809, 1883.

SCHLEBUSCH, Stephen; ILLING, Nicola. Next generation shotgun sequencing and the challenges of de novo genome assembly. *South African Journal of Science*, v. 108, n. 11–12, p. 62–70, jan. 2012.

SCHATZ, Michael C.; DELCHER, Arthur L.; SALZBERG, Steven L. Assembly of large genomes using second-generation sequencing. *Genome Research*, v. 20, n. 9, p. 1165–1173, set. 2010.

SHENDURE, J.; JI, H. Next-generation DNA sequencing. *Nature Biotechnology*, v. 26, n. 10, p. 1135–1145, out. 2008.

SILBER, A. M. *et al.* Active transport of L-proline in *Trypanosoma cruzi*. *The Journal of Eukaryotic Microbiology*, v. 49, n. 6, p. 441–446, dez. 2002.

SIMPSON, A. G. B. The identity and composition of the Euglenozoa. *Archiv für Protistenkunde*, v. 148, n. 3, p. 318–328, 1 out. 1997.

SIMPSON, A. G. B.; LUKES, J.; ROGER, A. J. The evolutionary history of kinetoplastids and their kinetoplasts. *Molecular Biology and Evolution*, v. 19, n. 12, p. 2071–2083, dez. 2002.

SIMPSON, A. G. B.; STEVENS, J. R.; LUKEŠ, J. The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology*, v. 22, n. 4, p. 168–174, 1 abr. 2006.

SINGH, Neeloo; CHIKARA, Surendra; SUNDAR, Shyam. SOLiD<sup>TM</sup> Sequencing of Genomes of Clinical Isolates of *Leishmania donovani* from India Confirm *Leptomonas* Co-Infection and Raise Some Key Questions. *PLOS ONE*, v. 8, n. 2, p. e55738, 13 fev. 2013.

SJÖLANDER, Kimmen. Phylogenomic inference of protein molecular function: advances and challenges. *Bioinformatics (Oxford, England)*, v. 20, n. 2, p. 170–179, 22 jan. 2004.

SONNHAMMER, Erik L. L.; KOONIN, Eugene V. Orthology, paralogy and proposed classification for paralog subtypes. *Trends in genetics: TIG*, v. 18, n. 12, p. 619–620, dez. 2002.

SOUSA, M. A. & CORTE-REAL, S. Postnuclear kinetoplast in choanomastigotes of *Crithidia deanei* Carvalho, 1973: a proposal of a new genus. *Rev Inst Med Trop*, v. 33, S8 1991.

STAMATAKIS, Alexandros. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, v. 30, n. 9, p. 1312–1313, 5 jan. 2014.

STEIN, L. Genome annotation: from sequence to biology. *Nature Reviews. Genetics*, v. 2, n. 7, p. 493–503, jul. 2001.

STEVENS, J. R. Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes. *Parasite (Paris, France)*, v. 15, n. 3, p. 226–232, set. 2008.

TEIXEIRA, M. M. G. *et al.* Phylogenetic Validation of the Genera *Angomonas* and *Strigomonas* of Trypanosomatids Harboring Bacterial Endosymbionts with the Description of New Species of Trypanosomatids and of Proteobacterial Symbionts. *Protist*, v. 162, n. 3, p. 503–524, 1 jul. 2011.

TAMAS, I. *et al.* 50 Million Years of Genomic Stasis in Endosymbiotic Bacteria. *Science*, v. 296, n. 5577, p. 2376–2379, 28 jun. 2002.

TAYLOR, M. J.; BANDI, C.; HOERAUF, A. Wolbachia. Bacterial Endosymbionts of Filarial Nematodes. In: BAKER, J. R.; MULLER, R.; ROLLINSON, D. (Org.). . *Advances in Parasitology*. [S.l.]: Academic Press, 2005. v. 60. p. 245–284. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0065308X05600048>>. Acesso em: 16 out. 2017.

TEIXEIRA, M. M. G. *et al.* Phylogenetic Validation of the Genera *Angomonas* and *Strigomonas* of Trypanosomatids Harboring Bacterial Endosymbionts with the Description of New Species of Trypanosomatids and of Proteobacterial Symbionts. *Protist*, v. 162, n. 3, p. 503–524, 1 jul. 2011.

TIBAYRENC, M.; AYALA, F. J. Evolutionary genetics of *Trypanosoma* and *Leishmania*. *Microbes and Infection*, v. 1, n. 6, p. 465–472, maio 1999.

VICKERMAN, K. The evolutionary expansion of the trypanosomatid flagellates. *International Journal for Parasitology*, v. 24, n. 8, p. 1317–1331, dez. 1994.

VOTÝPKA, J. *et al.* *Kentomonas* gen. n., a New Genus of Endosymbiont-containing Trypanosomatids of Strigomonadinae subfam. n. *Protist*, v. 165, n. 6, p. 825–838, 1 dez. 2014.

WALLIN, I. E. On the nature of mitochondria. IX. Demonstration of the bacterial nature of mitochondria. *American Journal of Anatomy*, v. 36, n. 1, p. 131–149, 1 set. 1925.

WERNEGREEN, Jennifer J. Endosymbiont evolution: Predictions from theory and surprises from genomes. *Annals of the New York Academy of Sciences*, v. 1360, n. 1, p. 16–35, dez. 2015.

YURCHENKO, V. *et al.* An Integrated Morphological and Molecular Approach to a New Species Description in the Trypanosomatidae: the Case of *Leptomonas podlipaevi* n. sp., a Parasite of *Boisea rubrolineata* (Hemiptera: Rhopalidae). *Journal of Eukaryotic Microbiology*, v. 53, n. 2, p. 103–111, 1 mar. 2006.

ZHANG, Jun *et al.* The impact of next-generation sequencing on genomics. *Journal of genetics and genomics = Yi chuan xue bao*, v. 38, n. 3, p. 95–109, 20 mar. 2011.