

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

Pedro Ammon Liszt Rocha de Moura Palumbo

Predição de estados ancestrais da histidina-amônia liase de kinetoplastidae

São Carlos

2022

Pedro Ammon Liszt Rocha de Moura Palumbo

Predição de estados ancestrais da histidina-amônia liase de kinetoplastidae

Trabalho de conclusão de curso apresentado ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Bacharel em Físicas Biomolecular.

Orientador: Prof. Dr. Otavio Thiemann - Instituto de Física de São Carlos

São Carlos

2022

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

RESUMO

Um conjunto de organismos que descende de um ancestral comum possui proteínas com uma mesma função, porém com sequências de aminoácidos que diferem entre si e que remetem a uma mesma proteína como ancestral, até mesmo com alguma adaptação em sua função. Em Kinetoplastidae a histidina amônia liase (HAL) é central no metabolismo de degradação de L-histidina convertendo a histidina em amônia e ácido urocânico. A HAL foi caracterizada em *Trypanosoma cruzi*, mas o gene codificante para a enzima não foi encontrado no genoma de *Leishmania sp.* Por ser uma enzima importante no metabolismo de histidina e sua ausência em alguns Kinetoplastidae, possibilita empregar a reconstrução de sequências ancestrais (ASR) para compreender o processo evolutivo dessa enzima e fornecer relações entre metabolismo e estilo de vida na família dos cinetoplastídeos. Desta forma a filogenia da HAL foi reconstruída em especial para as homólogas de organismos próximos ao ramo monofilético dos Kinetoplastidae. A sequência ancestral foi predita por máxima verossimilhança para, então, ser sintetizada para estudos estruturais e bioquímicos futuros. Desta forma, será possível determinar as características da enzima ancestral e derivar a sua história evolutiva.

Palavras-chave: *Trypanosoma cruzi*, histidine-ammonia lyase (HAL), Ancestral Sequence Reconstruction (ASR), Reconstrução de Sequências Ancestrais.

1 INTRODUÇÃO

1.1 Filogenia e ciclo de vida da *Trypanosoma cruzi*

Os protistas constituem um reino de organismos eucariontes. Dentro da classe *Kinetoplastea*, a família *Trypanosomatidae* inclui uma variedade diversa de espécies parasitárias que atuam como agentes de doenças significativas, como a Doença de Chagas (*Trypanosoma cruzi*), a doença do sono (*Trypanosoma brucei*) e as leishmanioses. Os tripanossomatídeos (família *Trypanosomatidae*) são organismos flagelados, caracterizados pela presença de uma única mitocôndria altamente modificada, o ciclo de vida complexo e algumas análises moleculares (de algumas proteínas essenciais) corroboram o grupo como sendo monofilético^[1,2]. Esses organismos evoluíram para habitar diferentes nichos ecológicos, tanto livremente quanto em invertebrados e vertebrados, até mesmo em plantas, como parasitas. As maiores características evolutivas desses seres são o desenvolvimento de cinetoplasto (acoplado à mitocôndria modificada), a perda de tRNAs próprios das mitocôndrias, compartimentos e mudanças de vias metabólicas entre outras mudanças para sobreviver^[2].

Entre os tripanossomatídeos, *T. cruzi*, o agente etiológico da Doença de Chagas, é especialmente interessante devido à sua complexidade metabólica e seu ciclo de vida típico de quatro estágios. Esse ciclo envolve fases em dois hospedeiros: um vertebrado (geralmente mamíferos, como hospedeiro) e um invertebrado (como insetos triatomíneos, vetor)^[2,3].

No hospedeiro vertebrado, o organismo assume diversas formas, das quais duas se destacam, uma forma intracelular chamada amastigota (forma arredondada, sem flagelo livre), onde se desenvolve até a transformação morfológica em tripomastigota (forma alongada com flagelo livre ao longo da célula e uma membrana ondulante desenvolvida), nessa fase do ciclo, ambas as formas replicativas populam a célula hospedeira por fissão binária, trocam material genético e algumas entram em estado de hibernação até que a célula lise. Quando esses tripomastigotas estão no sangue (e amastigotas nas células do sangue, como macrófagos) disseminados no organismo, os vetores podem ingerir o parasita durante a picada. No invertebrado, ele assume formas replicativas chamadas epimastigotas no intestino do inseto (forma alongada com flagelo emergindo próximo ao núcleo, parcialmente ligado ao corpo celular por uma membrana ondulante), sendo eliminado pelos excrementos [Figura 1]. A transição entre essas formas é crucial para sua sobrevivência em diferentes ambientes, como o sangue do hospedeiro vertebrado e o trato digestivo do vetor invertebrado, permitindo a disseminação do parasita^[2,3].

centrais no ciclo de vida do parasita, uma delas é objeto do presente estudo, a histidina-amônia liase presente nos acidocalcissomos.

1.2 Acidocalcissomo e histidina-amônia liase

O acidocalcissomo, uma organela altamente conservada em diversos organismos protozoários, desempenha um papel essencial no metabolismo de *T. cruzi*. Essa organela é rica em fosfatos (P_i , PP_i , PolyP) e atua na regulação da homeostase de cátions inorgânicos (Na^+ , K^+ , Mg^{2+} , Ca^{2+} , Fe^{2+} , Zn^{2+}) e orgânicos (arginina, lisina e poliaminas), armazenamento de cálcio e controle do balanço ácido-base, que depende da atividade de bombas de prótons dependentes de ATP e PP_i e da tonicidade dos meios, sendo essencial para a adaptação do parasita a ambientes variados, como o trato digestivo do inseto vetor e o sangue do hospedeiro vertebrado. A adição de histidina a epimastigotas permeabilizados ou acidocalcissomas isolados alcaliniza os acidocalcissomas, isso representa um mecanismo fisiológico de liberação de Ca^{2+} ligado ao metabolismo de aminoácidos [Figura 2A e 2B] [4,5].

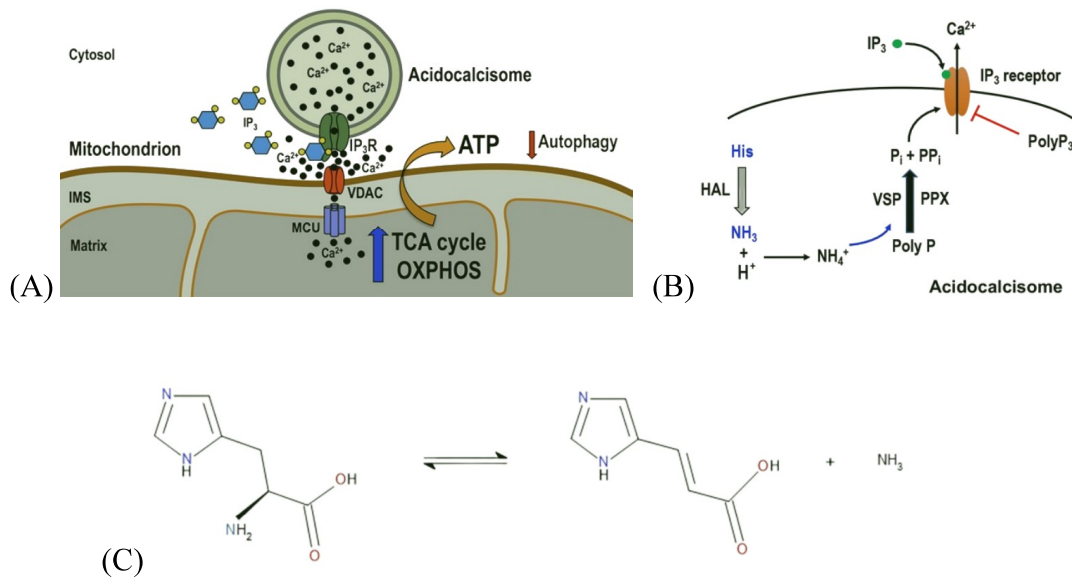


Figura 2. Esquema simplificado do acidocalcissomo e atividade da HAL. (A) O Ca^{2+} (pontos pretos) é liberado dos acidocalcissomos após a estimulação do receptor de inositol trisfosfato (IP_3R), o Ca^{2+} atravessa a membrana mitocondrial externa através do canal de ânion dependente de voltagem (VDAC), é processado pelo uniporte de Ca^{2+} mitocondrial (MCU) (B) A hidrólise do PolyP estimula a liberação de Ca^{2+} através do receptor de IP_3 dos acidocalcissomos. A desaminação da histidina (His) é catalisada pela histidina amônia-liase (HAL) para gerar amônia (NH_3) nos acidocalcissomos. (C) Reação catalisada pela histidase, uma histidina convertida em ácido urocânico e amônia. A e B adaptado de [5] e C de [9].

A enzima L-histidina-amônia liase (HAL EC:4.3.1.3), faz parte da família de liases de aminoácidos aromáticos, apresenta-se com estrutura homotetramérica, também conhecida como histidase, está localizada dentro do acidocalcisomo. Ela provoca alcalinização do ambiente ao converter histidina em amônia (NH_3) e ácido urocânico ($\text{C}_6\text{H}_6\text{N}_2\text{O}_2$) [Figura 2C] por esta reação, a etapa inicial da degradação da histidina, gerando ácido urocânico, posteriormente convertido em glutamato e demais vias de outros aminoácidos^[6]. Está metabolicamente ligada ao polifosfato (polyP) devido à região intrinsecamente desordenada nos resíduos ricos em lisina no C-terminal, isto é, quando o pH está baixo, os íons H^+ induzem a desafixação da HAL no polyP e promovem maior atividade da enzima por difusão na organela. Dessa forma, o produto gerado (NH_3) captura H^+ e aumenta o pH, favorecendo a atividade metabólica do organismo. É razoável hipotetizar que uma capacidade aumentada de desaminação de histidina seria prejudicial ao desenvolvimento intracelular na fase amastigota, de forma que a via da HAL ativa, principalmente, durante as formas no vetor epimastigota e tripomastigota ^[5].

A *Tc*HAL apresenta dois domínios, um central e outro catalítico, que foram identificados por sobreposição estrutural pelos pesquisadores^[7]. Além disso, também encontraram nas estruturas homólogas, com exceção de fungos ou plantas, carece do chamado domínio de proteção, uma região da proteína que protege o sítio ativo [Figura 3]. A superposição entre os quatro monômeros de *Tc*HAL não revelou diferenças estruturais significativas em suas conformações. Os aminoácidos catalíticos são conservados entre os seus homólogos, e seus resíduos foram identificados com base em estruturas homólogas resolvidas experimentalmente. As conformações estruturais das HAL são aproximadamente conservadas entre os organismos estudados por esses pesquisadores^[7]. A Cisteína na posição 393 da *Tc*HAL está na forma oxidada, mas em estruturas homólogas esse aminoácido é diferente e com exceções, que são as sequências das isoformas 1, 2 e 3 da HAL humana^[7].

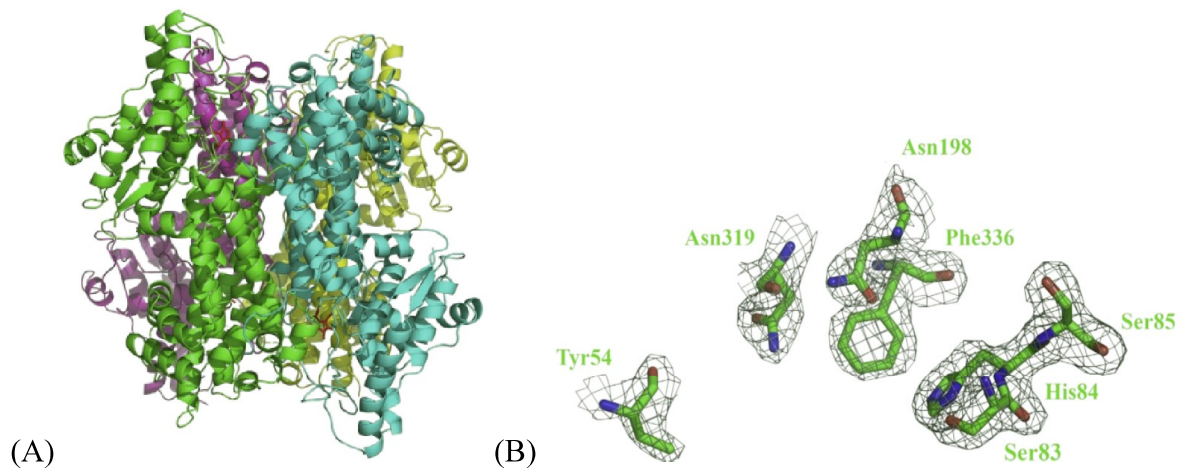


Figura 3. Estrutura da HAL. (A) Estrutura global do tetrâmero da *Tc*HAL. Cada monômero é representado por uma cor diferente e os cofatores MIO (3,5-diidro-5-metilideno-4H-imidazol-4-ona) por bastões vermelhos, estes são formados autocataliticamente pela ciclização e desidratação dos resíduos de alanina, serina e glicina, comum à família^[8]. (B) Qualidade do ajuste na densidade eletrônica para resíduos do sítio ativo e seletividade do substrato da *Tc*HAL, monômero C. Adaptado de [9].

Essas características estruturais e funcionais do acidocalcissomo e da *Tc*HAL apontam para adaptações evolutivas relevantes para a sobrevivência do parasita. Para entender melhor sobre o funcionamento e a razão evolutiva da enzima, a reconstrução de sequências ancestrais da HAL nos tripanossomatídeos permitirá aprofundar a compreensão sobre as modificações que diferenciam os tripanossomos das leishmanias.

1.3 Reconstrução de sequências ancestrais

A reconstrução de sequências ancestrais (Ancestral Sequence Reconstruction - ASR) é uma abordagem computacional e experimental utilizada para inferir as sequências genéticas ou proteicas de organismos ancestrais já extintos. Esta abordagem pode ajudar a esclarecer como essa enzima evoluiu nos trypanosomatídeos, revelando possíveis adaptações metabólicas ao longo de sua evolução, o papel de enzimas essenciais no ciclo de vida parasitário e nos mecanismos de sobrevivência. O princípio por trás da ASR é empregar os padrões de mutações dos genes e proteínas atuais, que evoluíram a partir de formas ancestrais, a deduzir as sequências de aminoácidos ancestrais. Este processo faz uso de modelos estatísticos evolutivos, ou seja, que consideram as taxas de mutação e os padrões de substituição de aminoácidos ao longo do tempo^[10].

Os métodos de ASR iniciam com a construção de uma árvore filogenética, que organiza as relações evolutivas entre as proteínas disponíveis a partir do alinhamento de suas estruturas primárias. Com o alinhamento, aplica-se a inferência probabilística, por meio de métodos como a máxima verossimilhança ou os métodos bayesianos, nos nós da árvore filogenética determinados pelos modelos evolutivos utilizados. É calculado, então, para cada posição da sequência proteica, os aminoácidos mais prováveis de compor a proteína ancestral, com base nas sequências modernas e nas relações evolutivas^[10]. Por fim, a modelagem estrutural da proteína pode ser feita pelo modelo de inteligência artificial AlphaFold^[20] ou utilizando técnicas experimentais de determinação da estrutura, como Cristalografia e difração de Raios-X de monocristais, Criomicroscopia Eletrônica ou Espectroscopia de Ressonância Magnética Nuclear (RMN), após a obtenção da proteína predita por expressão heteróloga.

Após a aplicação da ASR, é essencial aumentar a confiabilidade do modelo predito da proteína ancestral, podendo ser avaliada e comparada com base em suas características estruturais e biofísicas, como estabilidade térmica, atividade enzimática ou ensaios de ligação com cofatores.

Atualmente, é conhecida a estrutura experimentalmente da *Tc*HAL e o mecanismo de ação juntamente da via metabólica já são focos de estudos, porém não é conhecido uma enzima semelhante nas leishmanias, da mesma forma, não foi identificado os genes codificadores dela. Este estudo presta-se a identificar sequências, proteômicas e genômicas, e estruturas a partir da reconstrução de enzimas ancestrais dos tripanosomatídeos e utilizando a *Tc*HAL como modelo estrutural e funcional e também diversas sequências de aminoácidos para a reconstrução da ancestralidade.

2 MATERIAIS E MÉTODOS

A sequência de aminoácidos da proteína *Tc*HAL (GenBank PBJ76975.1 e UniProt Q4E133) foi obtida do banco de dados National Center for Biotechnology Information (NCBI). A ferramenta BLAST foi utilizada para identificar sequências homólogas, utilizando as configurações padrões do algoritmo, em especial, a matriz de substituição BLOSUM62 (BLOcks SUBstitution Matrix) foi utilizada^[11,12]. A matriz é desenvolvida atribuindo uma pontuação para cada mutação de um aminoácido para outro de acordo com a frequência observada em bloco de sequências de proteínas homólogas, refletindo a frequência de alterações evolutivas. Quanto maior o valor da pontuação de trocas entre aminoácidos (entre -4 e 11) mais provável de acontecer a mutação ou mais conservada tende a ser aquela região. O número que acompanha a seguir do nome da matriz (62, 45, 50, 80) é o limite de identidade (percentual) entre os blocos das estruturas primárias idênticas a ser considerado, ou seja, quanto menor, mais diferente tendem a ser os organismos relacionados^[12].

As sequências homólogas encontradas foram selecionadas com base nos seguintes critérios: cobertura da sequência (>90%), similaridade (>60%), *E value* próximo de 0. Também foram consideradas sequências pertencentes a classes e ordens dos organismos pertencentes a classe trypanosomatida (trypanosoma, leishmania, herpetomonas, strigomonas) e blastocrithidia, o organismo *Bodo saltans* ainda pertencente a classe kinetoplastida e de ordem bodonidae como sendo o organismo mais próximo fora dos kinetoplastidas^[6] e, como outgroup, foi coletado o organismo *Naegleria gruberi*, pertencente ao filo amoebozoa, por possuir uma HAL com mais de

50% de identidade. Além disso, para aprofundar melhor as diferenciações ocorridas, também foram coletados outros organismos drasticamente diferentes, pertencentes ao reino animalia, pois também possuem uma identidade de sequência maior que 50% com a *Tc*HAL e podem trazer informações sobre a convergência evolutiva para as sequências de leishmanias. Demais sequências da HAL de trypanosomas putativas não foram utilizadas, pois possuem o mesmo alicerce experimental das estruturas resolvidas e sequências genômicas catalogadas, tornando-as redundantes.

As sequências escolhidas foram então alinhadas empregando o software MEGA11^[13], utilizando o método MUSCLE (Multiple Sequence Comparison by Log-Expectation). O software MUSCLE emprega um algoritmo amplamente utilizado e conhecido por ser rápido e eficiente, especialmente quando se trata de grandes conjuntos de dados, combinando métodos de alinhamento globais e locais^[13]. O MUSCLE inicia criando uma árvore filogenética guia a partir das identidades de pares de sequências pelo método de neighbor-joining. Em seguida, realiza alinhamentos de pares baseados nessa árvore, gerando um alinhamento inicial, que é refinado por iterações, recalculando a árvore e ajustando as sequências até atingir um alinhamento em que não haja melhorias significativas^[14].

Após essa etapa, o software foi utilizado para inferir a árvore filogenética por máxima verossimilhança^[16] empregando o modelo evolutivo JTT (Jones-Taylor-Thornton) cujo é definido por uma matriz de substituição que fornece as taxas de transição entre todos os pares de aminoácidos, permitindo calcular as probabilidades de observar uma sequência de aminoácidos dada uma árvore filogenética e um tempo de evolução^[15]. Baseada na teoria de Markov, neste contexto, os aminoácidos são os estados e as substituições de um aminoácido são as transições, que podem ser reajustadas iterativamente^[16]. O método heurístico NNI (Nearest Neighbor Interchange) opera fazendo pequenas mudanças na estrutura da árvore em estudo, trocando pares de ramos adjacentes para explorar possíveis rearranjos que possam aumentar a verossimilhança baseada na matriz JTT a cada iteração^[17].

Para dar mais confiabilidade estatística à árvore gerada, o próprio software fornece a opção de bootstrap^[18], um técnica para estimar a precisão dos resultados, onde novas listas das sequências são geradas a partir da original pela seleção aleatória e com reposição até que tenha o mesmo número de elementos (i.e. algumas sequências podem ser escolhida várias vezes e outras nenhuma), então são geradas novas árvores e contabilizada a frequência que cada nó se mantém idêntico. Foram empregadas 700 interações de bootstrap e foram considerados nós confiáveis aqueles com mais de 70% de confiabilidade. O grupo externo (*outgroup*) escolhido foi a HAL de *Naegleria*

gruberi, que, apesar de não ser o organismo mais distinto do *T. cruzi*, é o intermediário entre os kinetoplastida e animalia.

Para verificar a robustez da árvore criada, o software GBlocks^[19] foi utilizado para após o alinhamentos múltiplos de sequências para identificar e remover regiões que não são suficientemente conservadas, aplicando critérios de gaps e a variabilidade entre as sequências e o ajuste de parâmetros, como a porcentagem mínima de sequências necessárias para conservar um bloco e o número máximo de gaps permitidos, para determinar quais blocos de sequência devem ser mantidos^[19]. Os seguintes parâmetros foram ajustados: Mínimo de sequências para posições conservadas: 7; Mínimo de sequências para uma posição flanqueadora: 11; Máximo de posições não conservadas contíguas: 8; Comprimento mínimo de um bloco: 5; Posições de gaps permitidos: Nenhuma; Matrizes de Similaridade: Sim. Então, uma nova árvore dessas sequências editadas é gerada pelos mesmos métodos, quando comparadas, devem ser estatisticamente idênticas.

O ancestral de trypanosoma corresponde ao nó 16 da árvore e o ancestral antes da diferenciação com o *Boto Saltans* está representado pelo nó 24 [Figura 5]. Ambos são importantes para a análise, pois representam uma diferenciação chave na filogenia dos kinetoplastideos, como na árvore apresentada pelo trabalho de Alexei Yu. Kostygov, Julius Lukeš et al [Figura 4].

O próprio software MEGA 11 fornece uma sequência precursora mais provável para cada nó que relaciona a ancestralidade dos organismos na árvore pelo método de máxima verossimilhança.

Para identificar semelhanças estruturais, as sequências ancestrais foram preditas pelo AlphaFold^[20], versão 3, e comparada com a *TcHAL*, de estrutura conhecida. A busca por semelhanças nos sítios ativos, *motifs* equivalentes pelas sequências, são verificados pelo alinhamento estrutural entre a *TcHAL* e os ancestrais citados anteriormente devem coincidir minimamente.

Pesquisadores das universidades brasileiras UEPG, UTFPR e USP foram pioneiros em obter a estrutura de uma histidase de eucariotos, eles constataram que a composição do sítio ativo é Tyr54, Asn198, Asn319 e Phe336 mais os resíduos Ser83, His84, Ser85 que estão relacionados a seletividade do substrato^[9]. Portanto, além da estrutura similar proporcionada pela evolução da enzima, outra pista importante para ser verificada são os aminoácidos presentes no sítio ativo. [Figura 3].

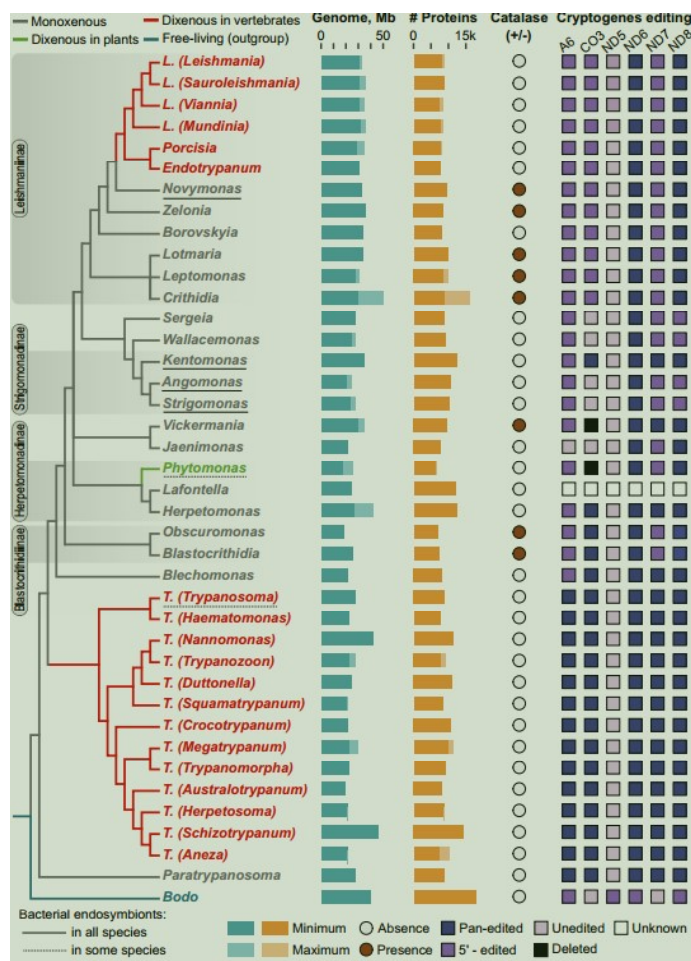


Figura 4. A árvore filogenética de Trypanosomatídeos mapeia características genômicas a partir de análises independentes. A árvore combina análises no nível de gênero para toda a família, e no nível de espécie para *Leishmaniinae* de ciclo simples, o gênero *Trypanosoma* e outra análise publicada anteriormente ao estudo para o clado relacionado. Adaptado de [4].

O DALI (Distance Matrix Alignment) é um dos métodos de alinhamento e comparação de estruturas tridimensionais empregado para o presente estudo. É baseado em uma medida sensível de semelhanças geométricas definida como uma soma ponderada das semelhanças das distâncias intramoleculares, dessa forma, reflete as similaridades estruturais, mesmo em sequências de aminoácidos não relacionada^[21], isso torna essa ferramenta muito informativa e complementar às análises que utilizam apenas a sequência de aminoácidos. O FATCAT (Flexible structure Alignment by Chaining Aligned fragment pairs allowing Twists) é, também, uma boa ferramenta para comparar estruturas, compara proteínas, otimizando o alinhamento e minimizando movimentos rígidos na estrutura. Suas funções permitem alinhar duas cadeias proteicas, buscar proteínas similares e comparar diferenças estruturais com homólogos próximos^[22].

3 RESULTADOS

Para o alinhamento de sequências homólogas, a coleta de sequências no BLAST resultou nas seguintes sequências: HAL *T. cruzi cruzi* (PBJ76975.1), HAL *T. cruzi Dm28c* (ESS65584.1), HAL *T. grayi* (XP_009307484.1), HAL *T. theileri* (XP_028879590.1), HAL *T. rangeli SC58* (ESL10714.1), AAL *T. melophagium* (KAH9577937.1), HAL *B. lanceolatum* (CAH1266363.1), HAL *P. maximus* (XP_033744655.1), HAL *C. fasciculata* (XP_004352426.1), HAL *T. rubripes* (XP_003967702.1), HAL *T. flavidus* (XP_056908589.1), HAL hipotética *B. saltans* (CUG93556.1). O uso de homólogos hipotéticos foi evitado, uma vez que os mesmos utilizam os homólogos não hipotéticos que serão empregados nesta análise. O *B. saltans* é exceção, por ser um organismos de referência entre os mais próximos relacionados dos bodonídeos em relação aos tripanosomatídeos, servirá como um modelo do ancestral dos tripanosomatídeos, permitindo a distinção entre as partes derivadas dos genomas dos parasitas (aquelas únicas dos tripanosomatídeos) e aquelas que são o essencial para viver uma vida livre.

Também foram feitas análises traduzidas do proteoma para o genoma desses seres próximos no BLAST, e os resultados mostrados também não revelaram nada além de semelhanças em outros trypanosomas.

Conforme a estrutura filogenética construída por Yu. Kostygovm, Alexei et al., os trypanosoma são os primeiros diferenciarem-se entre os tripanosomatídeos [Figura 4]^[4]. Eles se diferenciam a partir do *B. saltans* e *N. gruberi* (percolozoa, outro filo) e compõem o supergrupo excavata. Os demais organismos apresentados são mais aproximados aos animais, que conferem aos cálculos dos métodos da árvore informações sobre grandes evoluções no tempo. As HALs das espécies (5) e (3) de trypanosoma ficaram inconclusivas, ao comparar com o bootstrap [Figura 5A], provavelmente por falta de intermediários que tenham suas respectivas homólogas, cadastradas no servidor NCBI ou naturalmente não há. A árvore com as sequências após a edição pelo GBlocks apresentou uma proximidade maior entre essas mesmas espécies indica que o software excluiu sentenças importantes na diferenciação entre as mesmas [Figura 5B], sendo crucial para a adaptação ou não.

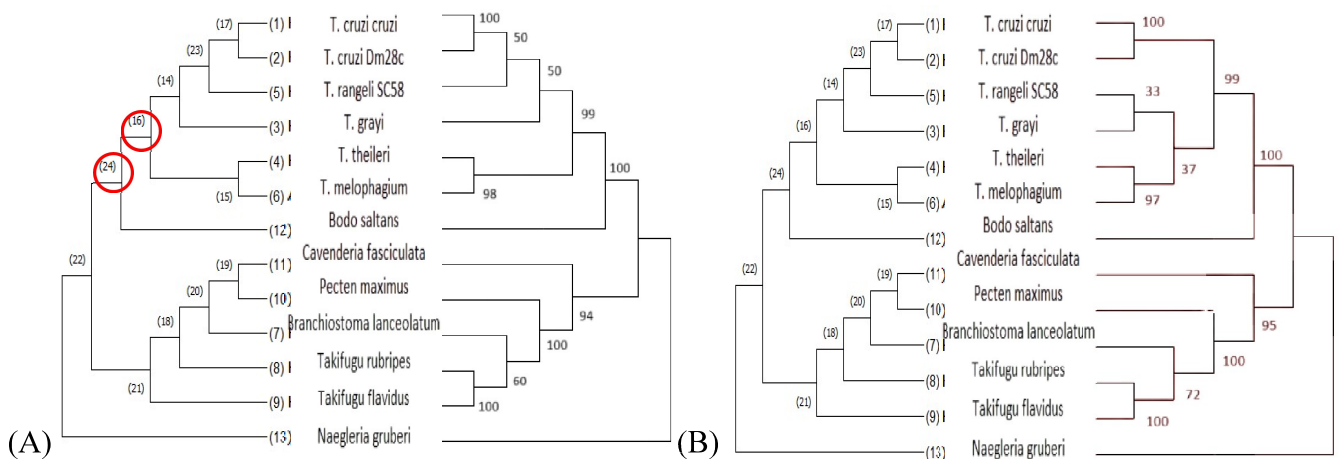


Figura 5. Árvores filogenéticas criadas pelo método de máxima verossimilhança (ML). O círculo vermelho destaca a posição dos ancestrais mais comuns entre os tripanosomas (nó 16) e entre o organismo mais próximo fora da ordem dos tripanosomatídeos. (A) Comparativo entre a árvore ML e o bootstrap (700 iterações) a partir da mesma. (B) Comparativo entre a mesma árvore por ML e a realizada a partir da edição em blocos do alinhamento (GBlocks)

As sequências dos nós 16 e 24 foram calculadas por máxima verossimilhança no MEGA11 (mesmo método para criar a árvore filogenética), este exporta o ancestral mais provável. Com estas sequências, o primeiro passo é realizar uma comparação entre si (Figura 6) e suas homólogas com um alinhamento global [Tabela 1], dessa forma, é informado quanto de identidade foi preservado entre elas dentre as HAL encontradas nos trypanosoma.

Nó 24	1	MPPVQSQILDGSSLTPDALYALGYENARILSEEARVAARKVIDDIVSRQVYVYGINTGFG	60
Nó 16	1	...M...V...C.....T.....I...A...K..E..T.....	60
Nó 24	61	KFESTIIPPQLEQLNLRSHSAGVGPLIPERARMMALRINVLAKGHSGRLEIVKLVVA	120
Nó 16	61C.....M...V...C.....Y..	120
Nó 24	121	FNAGVIPIPKGTVGASGDLAPLSHLALGMLGEGLLASNTFRDAAVLRELGVPEILAAKEG	180
Nó 16	121V...Q.....G.....T.K.....	180
Nó 24	181	LALINGTQFISALGAEAVVRARRIARLADVALAMTHEALSTVSTLDPDIHAVRPHKGQQL	240
Nó 16	181S.....N...R.....	240
Nó 24	241	VARLRSLHSEKYPGNEKKFEISEINSHNCGRVQDAYSIRCAPQVHGISNDMIEWVYGI	300
Nó 16	241A.....M.....L	300
Nó 24	301	TTELNCATDNPLVFPDGVGSQVPNSHHIHLDDENSRSSELHLLTTSLENDKIVSGGNFHGEY	360
Nó 16	301V.....	360
Nó 24	361	PAKALDMLAIGVHELGNISERRIERLNNPSLSRLPAFLVNGGLNSGFMIHCTAAALVSE	420
Nó 16	361	420
Nó 24	421	NKVI CHPASVDSTSTSAAQFDHVSMMGGFSARKAI KVVFMVFTTATEFI I CACQGTDI I RPI	480
Nó 16	421	...Y.....	480
Nó 24	481	KTTPLEKVVLRVSVSPWKDRVVRDIENVLLKSGKVVWVKIPEEARSVLTRKPFDSKSKL	538
Nó 16	481	...M.....D...R.....G.....EL....	538

Figura 6. Alinhamento global entre as sequências ancestrais para comparação pelo BLAST. Foram obtidas uma pontuação de 2590 e uma identidade de 94%, uma similaridade bem próxima entre ambos, como esperado.

Enzima alinhada	Identidade (%) com o nó 16	Identidade (%) com o nó 24
HAL ancestral comparadas entre si	94	94
TcHAL	81	75
TcDm28cHAL	68	64
TgHAL	78	74
TtHAL	80	74
TrSC58HAL	75	71

TmAAL	82	76
-------	----	----

Tabela 1. Comparação de alinhamentos globais entre as enzimas ancestrais e homólogas em trypanosomatidae.

Com as HALs ancestrais obtidas, é possível realizar a predição pelo algoritmo de rede neural AlphaFold. A partir das estruturas preditas, o próximo passo é a análise do alinhamento estrutural realizado pelo DALI e FATCAT.

Pela sobreposição das proteínas, nota-se que há uma grande semelhança entre elas, até mesmo no sítio ativo, com as posições relativas dos aminoácidos catalíticos ligeiramente diferentes, de forma que mantém a cavidade e a atividade semelhantes. Há uma região intrinsecamente desordenada perto do C-terminal, conforme a teoria apresentada por Barisón, M. J. et al, cuja fixação com a polyP é um importante regulador da atividade dentro do acidocalcisoma. Outra região desordenada no meio da cadeia polipeptídica, modelada pelo AlphaFold, não é encontrada na nativa de *T. cruzi*, o que pode significar uma má predição no local, além disso, essa região não é encontrada na *TcHAL*, possivelmente houve inserções ou deleções de aminoácidos não conclusivos sobre sua vantagem ou desvantagem evolutiva.

Como ambas as enzimas preditas são muito semelhantes, ou seja, bem sobrepostas, ambos os ancestrais resultaram em buscas semelhantes, provavelmente por haver uma grande sobreposição estrutural [Figura 6] e entre as sequências, derivada das nativas em tripanosomas.

A busca por homologia no FATCAT com a base de dados completa do PDB não produziu resultados significativos, apenas duas enzimas foram relevantes, uma delas é a própria *TcHAL* (com identidade de 63% de identidade para o ancestral do nó 24 e 65% para o nó 16. A outra é de *Pseudomonas putida*, de comprimento semelhante com identidade de ~35%, um desvio quadrático baixo (~2 Å). O organismo *P. putida* é muito distante dos tripanosomatídeos, pertencente ao domínio das bactérias, portanto, foi descartada para estudos relacionados à razão evolutiva em leishmanias.

No DALI, a pesquisa com todo o PDB, gerou resultados diversos, de maneira geral, nenhuma estrutura polipeptídica obteve identidade maior que 40%, com exceção da *TcHAL* (com 78% para o ancestral do nó 24 e 80% para o nó 16). Além disso, também houveram diversas enzimas com atividades catalíticas muito diferentes e outras liases de aminoácidos aromáticos (da mesma família de enzimas), essas não serão consideradas, pois, apesar de ter alguma similaridade,

possivelmente é pela similaridade do sítio ativo por catalisar aminoácidos parecidos, levando à outras adaptações das vias metabólicas fora do objetivo do presente estudo. As enzimas putativas também não foram consideradas, já que são informações ruidosas para esse tipo de análise.

Monômero	Enzima	Identidade (%) (nós 16 - 24)	RMSD (nós 16 - 24)	Organismo
1gk2-A	HAL	40 - 40	1.9 - 1.9	<i>Pseudomonas putida</i>
1b8f-A	HAL	39 - 40	1.9 - 2.0	<i>Pseudomonas putida</i>
1eb4-A	HAL	39 - 40	1.8 - 1.9	<i>Pseudomonas putida</i>
1gkj-A	HAL	39 - 40	2.0 - 2.0	<i>Pseudomonas putida</i>
1gkm-A	HAL	39 - 40	1.9 - 2.0	<i>Pseudomonas putida</i>
1gk3-A	HAL	40 - 40	1.9 - 2.0	<i>Pseudomonas putida</i>
6v6h-A	HAL	80 - 78	1.9 - 1.7	<i>Trypanosoma cruzi</i>
3kdy-A	HAL	37 - 37	2.5 - 2.2	<i>Streptomyces globisporus</i>

Tabela 2. Dados dos alinhamentos estruturais realizados pelo DALI. Comparação entre monômeros de HALs mais próximos na sobreposição para ambas as enzimas ancestrais. Coletado e adaptado dos resultados pelo DALI, autoria própria.

Considerando apenas as demais histidases, as similaridades não correspondem a nenhum organismo da família trypanosomatidae para nenhum dos ancestrais preditos, os 7 monômeros mais próximos em sua estrutura, ou seja, com maior identidade e menor desvio quadrático médio das posições dos resíduos, são de *Pseudomonas putida* e *Streptomyces globisporus*, conforme a Tabela 2.

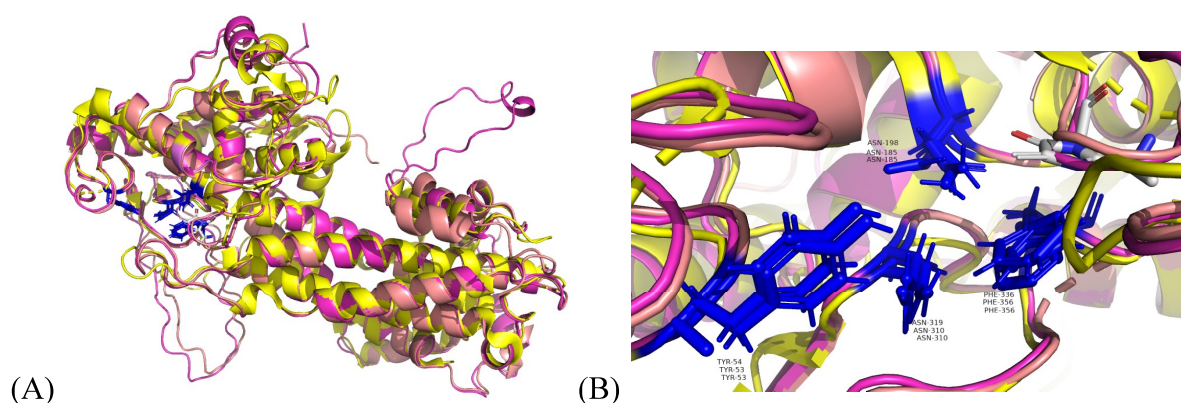


Figura 6. Alinhamento estrutural das proteínas ancestrais preditas com a TcHAL obtida experimentalmente. (A) estrutura predita do nó 16 (ancestrais dos tripanosomas) (rosa escuro) + estrutura predita do nó 24 (ancestrais dos tripanosomas) (rosa claro) + estrutura cristalizada da TcHAL alinhadas pelo FATCAT e adaptadas com o Pymol. (B) Aproximação da visualização do alinhamento no sítio ativo em azul.

4 CONCLUSÕES E CONSIDERAÇÕES FINAIS

A enzima histidina-amônia liase (HAL) tem recebido atenção no estudo devido à sua implicação no metabolismo dos aminoácidos. No entanto, ainda há desafios na compreensão do papel específico da HAL entre os tripanosomatídeos, e *Leishmania*, por exemplo, pode fornecer pistas importantes sobre adaptações ao longo da evolução que levaram à ausência desta enzima em seu metabolismo. Esse caso sugere que *Leishmania* desenvolveu estratégias metabólicas distintamente diferentes em comparação com outros organismos da classe, principalmente por ter sido se diferenciando a dixereno dentre os monoxenos, diferentemente do trypanosoma que se evoluíram de organismos não parasitas.

A análise da HAL e de outras enzimas liases de aminoácidos aromáticos em trypanosomas, embora essencial, enfrenta a limitação de modelos evolutivos que podem não reproduzir completamente as condições de mutação e seleção natural pelas quais os cinetoplastídeos passaram. Como o modelo de substituição JTT é empírico, ele se baseia em uma ampla diversidade de dados, mas pode apresentar limitações ao analisar organismos com estilos de vida especializados e trajetórias evolutivas como os dos cinetoplastídeos. Essa divergência pode interferir na simulação evolutiva, já que organismos com ciclos de vida complexos, como os tripanosomatídeos, podem ter sofrido mutações ao longo do tempo de modo distinto, por exemplo, pela transferência horizontal de genes.

Outro ponto de limitação para o algoritmo é a disponibilidade de sequências de HAL entre os tripanosomatídeos, o que pode dificultar uma análise estrutural robusta, aumentando as incertezas na reconstrução de sequências ancestrais. A inclusão de enzimas adicionais poderia fornecer mais dados e, potencialmente, uma reconstrução evolutiva mais detalhada. Em alguns casos, alternativas menos prováveis de reconstrução poderiam trazer à tona peculiaridades do ciclo de vida dos tripanosomatídeos, refletindo suas próprias trajetórias evolutivas.

A pesquisa sobre a HAL e outras enzimas correlatas entre os tripanosomatídeos e organismos relacionados têm muito a ser desenvolvida, e muitos aspectos bioquímicos, estruturais e evolutivos ainda permanecem pouco explorados.

REFERÊNCIAS

1. STEVENS, J.R. **Kinetoplastid phylogenetics, with special reference to the evolution of parasitic trypanosomes.** *Parasite*, v. 15, n. 3, p. 226-232, set. 2008. DOI: <https://doi.org/10.1051/parasite/2008153226>.
2. ZUMA, A.A.; GOMES, C.C.; RIBEIRO, D.; **Basic biology of *Trypanosoma cruzi*.** *Current Pharmaceutical Design*, dez. 2020. DOI: 10.2174/1381612826999201203213527.
3. MARTÍN-ESCOLANO, Javier et al. **An updated view of the *Trypanosoma cruzi* life cycle: intervention points for an effective treatment.** *ACS Infectious Diseases*, v. 8, n. 6, p. 1107-1115, 2022. DOI: 10.1021/acsinfecdis.2c00123.
4. KOSTYGOV, ALEXEI YU et al. **Phylogenetic framework to explore trait evolution in Trypanosomatidae.** *Trends in Parasitology*, v. 40, n. 2, fevereiro 2024. DOI: <https://doi.org/10.1016/j.pt.2023.11.009>.
5. DOCAMPO, Roberto; HUANG, Guozhong. **New insights into the role of acidocalcisomes in trypanosomatids.** *Journal of Eukaryotic Microbiology*, v. 67, n. 3, p. 375-388, 2020. DOI: 10.1111/jeu.12899. DOI: <https://doi.org/10.1111/jeu.12899>.
6. Barisón, M. J.; Damasceno, F. S.; Mantilla, B. S.; Silber, A. M. **Metabolomic profiling reveals a finely tuned, starvation-induced metabolic switch in *Trypanosoma cruzi* epimastigotes.** *Journal of Biological Chemistry*, v. 292, n. 21, p. 8964-8977, 2017.
7. Mantilla BS, Azevedo C, Denny PW, Saiardi A, Docampo R. 2021. **The histidine ammonia lyase of *Trypanosoma cruzi* is involved in acidocalcisome alkalization and is essential for survival under starvation conditions.** *mBio* 12:e01981-21. DOI: <https://doi.org/10.1128/mBio.01981-21>.

8. Schwede, T. F.; Rétey, J.; Schulz, G. E. **Crystal structure of histidine ammonia-lyase revealing a novel polypeptide modification as the catalytic electrophile.** *Biochemistry*, v. 38, n. 17, p. 5355-5361, 1999. Disponível em: <https://doi.org/10.1021/bi982929q>.
9. MIRANDA, ROBSON R.; SILVA, MARCIO; BARISON, MARIA J.; SILBER, ARIEL M.; IULEK, JORGE. **Crystal structure of histidine ammonia-lyase from *Trypanosoma cruzi*.** *Biochimica et Biophysica Acta*, 2020. Aceito em 17 mai. 2020. DOI: <https://doi.org/10.1016/j.biochi.2020.05.009>.
10. LIBERLES, David A. **Ancestral sequence reconstruction.** 1. ed. New York: Oxford University Press, 2007. ISBN 978-0-19-929232-8.
11. ALTSCHUL, Stephen F. et al. **Basic local alignment search tool.** *Journal of Molecular Biology*, v. 215, n. 3, p. 403-410, 1990. DOI: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
12. HENIKOFF, S.; HENIKOFF, J. G. **Amino acid substitution matrices from protein blocks.** *Proceedings of the National Academy of Sciences*, v. 89, n. 22, p. 10915-10919, 1992. DOI: <https://doi.org/10.1073/pnas.89.22.10915>.
13. TAMURA, K.; STECHER, G.; KUMAR, S. **MEGA11: Molecular Evolutionary Genetics Analysis version 11.** *Molecular Biology and Evolution*, v. 38, n. 7, p. 3022-3027, 2021. DOI: <https://doi.org/10.1093/molbev/msab120>.
14. EDGAR, R. C. **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Research*, v. 32, n. 5, p. 1792-1797, mar. 2004. DOI: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340). DOI: <https://doi.org/10.1093/nar/gkh340>.
15. JONES, David T.; TAYLOR, William R.; THORNTON, Janet M. **The rapid generation of mutation data matrices from protein sequences.** *Computers & Chemistry*, v. 16, n. 3, p. 275-282, 1992.

16. BYKOVA, Nadezda A.; FAVOROV, Alexander V.; MIRONOV, Andrey A. **Hidden Markov Models for Evolution and Comparative Genomics Analysis**. PLoS ONE, v. 8, n. 6, e65012, 2013. DOI: 10.1371/journal.pone.0065012. DOI: <https://doi.org/10.1371/journal.pone.0065012>.
17. FISCHER, Mareike. **On the correctness of maximum parsimony for data with few substitutions in the NNI neighborhood of phylogenetic trees**. arXiv:2403.01282. 2024. Acesso em: arXiv:2403.01282.
18. WOLLENBERG, Kurt R.; ATCHLEY, William R. **Separation of phylogenetic and functional associations in biological sequences by using the parametric bootstrap**. Proceedings of the National Academy of Sciences, v. 97, n. 5, p. 2474-2479, 2000. DOI: 10.1073/pnas.030540497.
19. CASTRESANA, J. **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis**. *Molecular Biology and Evolution*, v. 17, n. 4, p. 540-52, 2000
20. JUMPER, John et al. **Highly accurate protein structure prediction with AlphaFold**. *Nature*, v. 596, n. 7873, p. 583-589, 2021. DOI: 10.1038/s41586-021-03819-2.
21. HOLM, L. **Using Dali for Protein Structure Comparison**. In: GÁSPÁRI, Z. (Ed.). Structural Bioinformatics. Methods in Molecular Biology, v. 2112. New York, NY: Humana, 2020. p. 29-42.
22. LI, ZHANWEN; JAROSZEWSKI, LUKASZ; IYER, MALLIKA; SEDOVA, MAYYA; GODZIK, ADAM. **FATCAT 2.0: TOWARDS A BETTER UNDERSTANDING OF THE STRUCTURAL DIVERSITY OF PROTEINS**. *NUCLEIC ACIDS RESEARCH*, v. 48, n. W1, p. W60–W64, 2020. DISPONÍVEL EM: [HTTPS://DOI.ORG/10.1093/NAR/GKAA443](https://doi.org/10.1093/nar/gkaa443).