

ARTIGO ORIGINAL

DESENHO *IN SILICO* DE SGRNAS E DONOR DNAs PARA DELEÇÃO VIA CRISPR/CAS9 DE ENZIMAS RELACIONADAS AO FENÓTIPO DE RESISTÊNCIA DO *TRYPANOSOMA CRUZI* AO BENZONIDAZOL

IN SILICO* DESIGN OF SGRNAS AND DONOR DNAs TO KNOCKOUT ENZYMES RELATED TO BENZNIDAZOLE-RESISTANCE PHENOTYPE IN *TRYPANOSOMA CRUZI

Isabella Fernandes Martins Santos^{1*}; Daniela de Melo Resende²; Silvane Maria Fonseca Murta³; Ana Maria Murta Santi⁴

1. Discente do curso de Biomedicina do Centro Universitário de Belo Horizonte – UniBH, 2020. Belo Horizonte, MG. <https://orcid.org/0000-0002-7970-0151isabellafmsantos@outlook.com>.
2. Doutora em Bioquímica e Imunologia. Universidade Federal de Minas Gerais, 2008. Instituto René Rachou. Belo Horizonte, MG. <http://orcid.org/0000-0003-2979-7984> daniela.resende@fiocruz.br
3. Doutora em Bioquímica. Universidade Federal de Minas Gerais, 1999. Instituto René Rachou. Belo Horizonte, MG. <https://orcid.org/0000-0002-8523-2155> silvane.murta@fiocruz.br
4. Mestre em Ciências da Saúde. Instituto René Rachou, 2017. Instituto René Rachou. Belo Horizonte, MG. <https://orcid.org/0000-0002-9791-6554> anamms552@gmail.com

* autor para correspondência: Isabella Fernandes Martins Santos: isabellafmsantos@outlook.com

Recebido em: 09/04/2021 - Aprovado em: 09/12/2021 - Disponibilizado em: 31/12/2021

RESUMO: A doença de Chagas, causada pelo protozoário *Trypanosoma cruzi*, afeta milhões de pessoas, principalmente na América Latina. Seu tratamento é baseado no uso clínico do benzonidazol (BZ) ou nifurtimox e possui limitações, como baixa eficácia de cura principalmente na fase crônica da doença e ocorrência de parasitos resistentes a ambos compostos. Várias enzimas têm sido relacionadas com o fenótipo de resistência dos parasitos a esses fármacos, tais como a nitroreductase 1, a aldo-keto reductase e a álcool desidrogenase. A fim de melhor compreender o papel dessas enzimas no fenótipo de resistência do *T. cruzi* ao BZ, propomos utilizar o CRISPR/Cas9 para realizar a deleção dessas enzimas. Foram realizadas análises *in silico* para o desenho dos sgRNAs e donor DNAs por meio da avaliação do número de cópias de cada gene na cepa CL Brener usando o tBLASTn, a construção dos sgRNAs e dos donor DNAs pela ferramenta EuPaGDT e o alinhamento global das cópias pelo Needle (EMBOSS). Foram encontradas duas sequências do gene *NTR-1*, duas para a *ADH* e pelo menos nove para a *AKR*,

sendo escolhidos três sgRNAs para reconhecimento específico de cada uma das cópias. Devido às características dos sgRNAs selecionados e eficiência do protocolo de CRISPR/Cas9, acreditamos que será possível realizar a deleção dos genes *NTR-1* e *ADH* na cepa CL Brener de *T. cruzi* utilizando essa metodologia caso não sejam genes essenciais para o parasito. Já no caso da *AKR* mais análises serão necessárias para conhecer o número exato de cópias do gene.

PALAVRAS-CHAVE: *Trypanosoma cruzi*. Resistência ao benznidazol. CRISPR/Cas9. Deleção gênica.

ABSTRACT: Chagas disease, caused by the protozoa *Trypanosoma cruzi*, affects millions of people, mainly in Latin America. Its treatment is based on clinical use of benznidazole (BZ) or nifurtimox and it has limitations, such as low cure efficacy, mainly in the chronic phase of the disease, and the occurrence of resistant parasites to both compounds. Several enzymes have been related to drug resistance phenotype in this parasite, such as Nitroreductase 1, aldo-keto reductase and alcohol dehydrogenase. In order to better understand the role of these enzymes in the drug resistance phenotype in *T. cruzi*, in the present work we propose to use the CRISPR/Cas9 to perform the knockout of genes encoding these enzymes. In this work, *in silico* analyzes were performed for the design of sgRNAs and donor DNAs by evaluating the number of copies of each gene in CL Brener strain using tBLASTn, the construction of sgRNAs and donor DNAs by EuPaGDT tool and the global alignment of copies by Needle (EMBOSS). Two sequences of the *NTR-1* gene were found, two for *ADH* and at least nine for *AKR*. Three sgRNAs were chosen for specific recognition of each enzyme. Due to the characteristics of the selected sgRNAs and the efficiency of the CRISPR/Cas9 protocol, we believe that it will be possible to perform the knockout of the *NTR-1* and *ADH* genes in the CL Brener *T. cruzi* strain using this methodology if the enzymes are not essential genes for the parasite. In the case of *AKR*, further analysis will be done to better evaluate the exact gene copy number.

KEYWORDS: *Trypanosoma cruzi*. Drug resistance. CRISPR/Cas9. Knockout.

1. INTRODUÇÃO

O *Trypanosoma cruzi* é um protozoário parasito hemoflagelado, agente etiológico da doença de Chagas (CHAGAS, 1909), que atualmente infecta cerca de 8 milhões de pessoas nas Américas (PEREIRA *et al.*, 2011; PÉREZ-MOLINA e MOLINA, 2017). Como muitos tripanosomatídeos, o *T. cruzi* possui um ciclo de vida heteroxênico, com hospedeiros triatomíneos hematófagos e hospedeiros vertebrados endotérmicos, incluindo humanos (GROOM; PROTOPAPAS; ZOCHIOS, 2017). A doença de Chagas se caracteriza por duas fases clínicas, uma aguda e a outra crônica. A fase aguda geralmente é assintomática sendo que, nessa fase, o paciente possui maior chance de cura após o tratamento, enquanto na fase crônica pode apresentar manifestações cardíacas, digestivas ou neurológicas (SARMENTO, 2008; URBINA e DOCAMPO, 2003; LOURDES, 2013; ARGOLO *et al.*, 2008; COURA, 2003).

Os fármacos utilizados para o tratamento da doença de Chagas são os compostos nitroheterocíclicos Nifurtimox (NFX) (Lampit®) e Benznidazol (BZ) (Rochagan®). Esses fármacos possuem eficácia limitada e a cura depende da susceptibilidade da cepa,

do estágio da doença e do estado imunológico do hospedeiro (URBINA e DOCAMPO, 2003; GARCIA-SALCEDO *et al.*, 2016). Na maioria dos casos, a doença só é detectada na fase crônica, o que dificulta o tratamento, já que os medicamentos se tornam menos eficazes nessa (GARCIA-SALCEDO *et al.*, 2016; ANDRADE *et al.*, 2008). Algumas cepas de *T. cruzi* possuem resistência natural a esses medicamentos, possuindo diversos mecanismos de adquirir essa resistência, o que pode explicar a falha no tratamento da doença (FILARDI e BRENER, 1987; CAMPOS *et al.*, 2014). Nesse contexto, a busca por novos alvos terapêuticos e conseqüentemente novos fármacos para tratamento da doença de Chagas se torna necessária.

O BZ e o NFX são pró-drogas que necessitam ser ativadas por enzimas do parasito, como por exemplo a nitroreductase do tipo I (NTR-1) ou prostaglandina F2alfa sintase (PGFS) (MURTA *et al.*, 2006; WILKINSON *et al.*, 2008; MEJIA *et al.*, 2012). Foi demonstrado que linhagens do *T. cruzi* que tiveram resistência induzida *in vitro* ao BZ apresentaram perda de cópias de uma dessas nitroreductases, seja da NTR-1 ou PGFS (MURTA *et al.*, 2006; WILKINSON *et al.*, 2008; MEJIA

et al., 2012). Essas enzimas realizam uma série de reduções do grupo nitro das pró-drogas BZ e NFX, cujos produtos irão promover danos ao DNA via estresse oxidativo (MCCALLA; REUVERS; KAISER, 1971; STREETER e HOENER, 1988; WILKINSON e KELLY, 2009; TROCHINE *et al.*, 2014).

Enquanto que por um lado o papel da PGFS na ativação dessas pró-drogas parece controverso (PETRAVICIUS *et al.*, 2019), os trabalhos da literatura parecem concordar que a NTR-1 é a principal enzima a realizar essa função em *T. cruzi*, utilizando o NADH como agente redutor. Essa enzima é insensível ao oxigênio, possuindo o mononucleotídeo de flavina (FMN) como cofator (WILKINSON *et al.*, 2008, PETRAVICIUS *et al.*, 2019).

Linhagens de *T. cruzi* resistentes foram associadas à perda de um cromossomo que continha o gene *TcNTR-1* e mutações na região de ligação do FMN do *TcNTR-1* tornaram a enzima inativa (WILKINSON *et al.*, 2008; MEJIA *et al.*, 2012). Wilkinson *et al.*, (2008) também demonstraram, através da deleção do gene *TcNTR-1* na cepa Sylvio- X10/6 de *T. cruzi*, que essa é uma enzima de extrema importância para a diferenciação do parasito, pois sem ela os parasitos não foram capazes de infectar células de mamíferos. A redução da expressão dessa enzima também tornou os parasitos mais resistentes ao BZ e ao NFX, demonstrando que essa enzima está relacionada com a resistência a esses compostos.

Outro estudo realizado por Campos *et al.*, (2014) demonstrou que parasitos resistentes ao BZ e com resistência cruzada ao NFX possuíam um códon de parada interrompendo o gene *TcNTR*, gerando uma proteína truncada. Interessantemente, outro estudo demonstrou que parasitos naturalmente resistentes ao BZ não apresentaram alterações na expressão da NTR-1, mas sim um aumento da expressão das enzimas aldo-keto redutase (AKR) e álcool desidrogenase (ADH) (GONZÁLES *et al.*, 2017). A AKR

é uma enzima NADPH-dependente que realiza a redução de naftoquinonas com efeito tripanocida (GARAVAGLIA *et al.*, 2010) e que em *T. cruzi* não está envolvida na biotivação do BZ, mas está relacionada à resistência do parasito a essa droga, provavelmente por atuar na remoção do glioxal (ROBERTS *et al.*, 2018). Por outro lado, a ADH faz parte de uma classe de oxiredutases que catalisam a oxidação reversível do etanol para o acetoaldeído, sincrônico à redução do NAD (REID e FEWSON, 1994). O papel da ADH na resistência dos parasitos ao BZ ainda não está completamente elucidado, uma vez que o proteoma da cepa DA de *T. cruzi* (naturalmente resistente ao BZ) demonstrou um aumento da expressão dessa enzima (GONZÁLES *et al.*, 2017), mas em *T. cruzi* 17LER, que possui resistência induzida *in vitro*, foi demonstrada uma redução dos transcritos dessa enzima (CAMPOS *et al.*, 2009).

Tendo em vista a importância da doença Chagas para saúde pública e a ocorrência de cepas de *T. cruzi* resistentes ao BZ, este artigo descreverá o desenho de RNAs guias (sgRNAs) e *donor DNAs* para realizar a deleção dos genes que codificam as enzimas NTR-1, AKR e ADH na cepa CL Brener utilizando o sistema CRISPR/Cas9. O sistema CRISPR é um mecanismo adaptável do sistema imune de bactérias que as protege de ácidos nucleicos virais (ISHINO *et al.*, 1987; JANSEN *et al.*, 2002; BOLOTIN *et al.*, 2005; GASIUNAS *et al.*, 2012; GARNEAU *et al.*, 2010; WIEDENHEFT, *et al.*, 2012; HORVATH e BARRANGOU, 2010). Nos últimos anos esse sistema vem sendo aprimorado e utilizado para edição de genomas de diversos organismos, com alta eficiência (JINEK, 2012; GASIUNAS *et al.*, 2012), incluindo o *T. cruzi* (LANDER *et al.*, 2016). O sistema CRISPR/Cas9 consiste em uma endonuclease (Cas9) e um RNA de guia único (sgRNA) que direciona a Cas9 para o DNA alvo que será clivado. Esse direcionamento é dado através de 20 nucleotídeos presentes na sequência do sgRNA (*protospacer*) que reconhecem uma sequência

complementar no genoma alvo. O DNA alvo deve estar localizado imediatamente a um *protospacer adjacent motif* (PAM), necessário para o reconhecimento da Cas9 (DELTCHEVA *et al.*, 2011; HSU *et al.*, 2013; SANDER e JOUNG, 2014). Na maioria das espécies, o reparo da dupla fita é realizado pela união de extremidade não homóloga (NHEJ), enquanto no *T. cruzi* ocorre exclusivamente pela junção não-homóloga alternativa (MMEJ) na ausência de um *template* (*donor DNA*). Por outro lado, na presença do *donor DNA* o reparo ocorre por recombinação homóloga, inserindo-se a sequência de interesse no local da quebra (PENG *et al.*, 2014; LANDER; CHIURILLO; DOCAMPO, 2016). Ao utilizar o sistema CRISPR/Cas9, deve-se observar a presença de *on-targets* e *off-targets*, pois essas sequências podem interferir no resultado final da edição do gene. *Off-targets* são sítios genômicos não intencionais nos quais a Cas9 pode se ligar para realizar a clivagem (ZHANG *et al.*, 2015; NAEEM *et al.*, 2020) e que pode causar danos severos ao genoma do organismo como largas deleções e rearranjo genômico (CAI *et al.*, 2018).

O objetivo do presente trabalho foi realizar análises *in silico* para desenho dos sgRNAs e *donor DNAs* para deleção, via sistema CRISPR/Cas9, das enzimas NTR-1, AKR e ADH em *T. cruzi* CL Brener, incluindo a análise do número de cópias, a avaliação das regiões conservadas entre as cópias e a escolha dos sgRNAs e *donor DNAs*.

2. METODOLOGIA

2.1. BUSCA DE CÓPIAS POR ORTOLOGIA E SINTENIA

A avaliação do número de cópias dos genes NTR-1, AKR e ADH da cepa CL Brener de *T. cruzi* foi realizada usando o banco de dados de tripanossomatídeos TriTrypDB, versão 49 beta (disponível em <https://tritrypdb.org/tritrypdb/>) (ASLETT *et al.*, 2009)

utilizando os números de acesso TcCLB.506791.70, TcCLB.511287.49 e TcCLB.506357.50. Foi feita uma busca na Tabela de Ortologia e Sintenia tanto por sequências de *T. cruzi* CL Brener quanto de CL Brener Esmeraldo-like e CL Brener Non-Esmeraldo-like.

2.2. BUSCA DE CÓPIAS POR SIMILARIDADE DE SEQUÊNCIAS

Foi realizado o alinhamento local das cópias de cada um dos genes usando o programa tBLASTn disponibilizado no banco de dados de tripanossomatídeos TriTrypDB (disponível em <https://tritrypdb.org/tritrypdb/app/search/genomic-sequence/SequencesBySimilarity>) (ASLETT *et al.*, 2009). Para o tBLASTn, as sequências proteicas da NTR-1, da AKR e da ADH foram usadas como sequências de busca (*query*) contra um banco de dados de nucleotídeos (*subject*), sendo escolhidos para tal os genomas de CL Brener Esmeraldo-like e CL Brener Non-Esmeraldo-like. Nessa busca foi usado um e-value de 10^{-4} , o filtro de baixa complexidade foi mantido desligado e foi determinado que 50 seria o número máximo de alinhamentos retornados.

2.3. DESENHO DOS SGRNAS E DOS DONOR DNAs

Os RNA guias (sgRNAs) e *donor DNAs* para deleção da NTR-1, da AKR e da ADH foram desenhados usando a ferramenta EuPaGDT (disponível em <http://gma.ctegd.uga.edu/>) (PENG e TARLETON, 2015) usando como referência as sequências de NTR-1, AKR e ADH identificadas nas buscas realizadas no TriTrypDB. Para esse desenho foi escolhida a endonuclease *SpCas9* e a busca foi realizada nos genomas de *T. cruzi* CLBrener TriTrypDB-29 (*hybrid diploid, including unassigned contigs*) e de *T. cruzi* CLBrener TriTrypDB-32 (*hybrid diploid*). Foi

determinado que os braços de homologia do *donor DNA* teriam 30 pares de bases (pb) e que a sequência de nucleotídeos TAGATAGATAGctcgag, contendo códons de parada em diferentes fases de leitura (em letras maiúsculas) e um sítio de restrição para a enzima *XhoI* (em letras minúsculas), seria inserida após o reparo. O filtro de busca por micro-homologia foi mantido ligado para o desenho dos sgRNAs.

O desenho dos iniciadores para produção dos sgRNA *in vitro* foi realizado de acordo com instruções do fabricante do kit *TranscriptAid T7 High Yield Transcription Kit* (Thermo Fischer Scientific).

2.4. ALINHAMENTO GLOBAL - ALINHAMENTO DE SEQUÊNCIAS PAR A PAR

O Needle (EMBOSS) (disponível em https://www.ebi.ac.uk/Tools/psa/emboss_needle/) foi utilizado para realizar o alinhamento global das sequências genômicas de NTR-1, AKR e ADH encontradas para CL Brener, Esmeraldo-like e CL Brener Non-Esmeraldo-like.

3. RESULTADOS

3.1. CÓPIAS ENCONTRADAS NA BUSCA POR ORTOLOGIA E SINTENIA

A cepa CL Brener do *T. cruzi* é uma cepa híbrida e o seu genoma diploide é formado tanto pelos haplótipos Esmeraldo-like quanto pelo Non-Esmeraldo-like. Sendo assim, é necessário buscar as sequências em ambos haplótipos. A Tabela 1 mostra as sequências já anotadas para NTR-1, AKR e ADH que podem ser encontradas na busca por ortologia e sintenia do TriTrypDB para ambas cepas CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like. Nesta busca foram encontradas duas sequências para cada gene analisado.

3.2. NÚMERO DE CÓPIAS ENCONTRADAS NO ALINHAMENTO LOCAL

Neste alinhamento foram utilizadas as sequências proteicas da NTR-1, da AKR e da ADH da cepa CL Brener Non-Esmeraldo-like como *query* e os genomas das cepas CL Brener Esmeraldo-like e CL Brener Non-Esmeraldo-like como *subject*. O resultado foi obtido por meio da busca utilizando tBLASTn, que realizou o alinhamento local das sequências contra o genoma das cepas CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like. Como resultado foram obtidas duas sequências para NTR-1 (Fig. S1), que são as mesmas que já estão anotadas como NTR-1 no TriTrypDB. Um resultado semelhante foi obtido para ADH (TcCLB.511277.60), para a qual os resultados encontrados pelo tBLASTn também foram os mesmos daqueles já anotados para o gene. A cobertura para esse alinhamento foi de 100% (Fig. S2).

Ao comparar as sequências anotadas para a AKR, percebeu-se grande diferença entre as sequências TcCLB.511287.49 e TcCLB.511627.120, diferença que pode ser observada na Fig. S3, na qual o alinhamento global das sequências de aminoácidos demonstrou uma identidade de 28% e uma similaridade de 44,6%. Dessa forma, foi realizado o tBLASTn de cada uma dessas sequências de forma independente, sendo denominadas AKR49 (TcCLB.511287.49) e AKR120 (TcCLB.511627.120). Para a sequência da AKR49 foram retornados seis alinhamentos (*matches*) (Fig. S4), resultado diferente do indicado na Tabela de Ortologia e Sintenia. Entre os *matches* retornados pelo tBLASTn estão a própria AKR49 (TcCLB.511287.49), a AKR120 (TcCLB.511627.120), a TcCLB.505183.120 e a TcCLB.507801.114, além da sequência de um pseudogene (TcCLB.506213.50) e de uma sequência genômica para a qual nenhum transcrito está associado. Para a busca com a sequência da AKR120 foram retornados os mesmos seis *matches* (Fig. S5). No entanto, a cobertura observada para AKR49 foi de

100% no primeiro alinhamento e 88,6% no segundo alinhamento, sendo mais baixa para os alinhamentos subsequentes. De forma semelhante, os alinhamentos obtidos para AKR120 tiveram mais de 81% de cobertura nos quatro primeiros alinhamentos, sendo mais baixa nos dois últimos.

3.3. DESENHO DOS SGRNAS E DONOR DNAs

Para escolha dos sgRNAs e *donor DNAs* para deleção da NTR-1, da AKR e da ADH na cepa CL Brener de *T. cruzi*, foi utilizada a ferramenta EuPaGDT. Para cada gene foram selecionados os três sgRNAs de maior escores, que não apresentassem *off-targets* e que não estivessem nem muito no início nem muito no final da sequência codificante. A Tabela 2 elenca os sgRNAs escolhidos, com os números de identificação, sequência e escore dos mesmos. As predições do EuPaGDT indicam que os três sgRNAs selecionados não possuem *off-targets* e são capazes de reconhecer com sucesso ambas sequências do gene da NTR-1: tanto a sequência CL Brener Esmeraldo-like quanto a CL Brener Non-Esmeraldo-like (Figura S6). Um resultado muito semelhante foi observado para ADH (Figura S7). Nesses casos não foram observadas diferenças nos resultados quando a busca foi realizada nos genomas de *T. cruzi* CL Brener TritypDB-29 (*hybrid diploid, including unassigned contigs*) e de *T. cruzi* CL Brener TritypDB-32 (*hybrid diploid*).

Para as buscas da AKR, considerando as diferenças entre as sequências (Fig. S3) e que dificilmente um mesmo sgRNA reconhecerá ambas sequências, também foram realizadas buscas independentes usando as sequências da AKR49 e AKR120, semelhante ao que foi feito nas análises de busca por similaridade de sequências. Em ambos casos as buscas retornaram sgRNAs com apenas 2 *on-targets* caso fosse escolhido o genoma *T. cruzi* CL Brener TritypDB-32 (*hybrid diploid*). No caso da AKR49, os sgRNAs reconhecem de forma perfeita as sequências

TcCLB.511287.49 e TcCLB.506213.50. No caso da AKR120 os sgRNAs reconhecem de forma perfeita as sequências TcCLB.511627.120 e TcCLB.507801.114. Por outro lado, quando são levados em consideração os *contigs* não ordenados e orientados (genoma *T. cruzi* CL Brener TritypDB-29 *hybrid diploid, including unassigned contigs*) cada sgRNA apresentou de quatro a seis *on-targets* para cada uma das sequências de AKR testadas, reconhecendo também sequências presentes nas regiões não ordenadas do genoma, além das sequências que foram reconhecidas ao usar a versão 32 do genoma. Dessa forma, considerando em conjunto as sequências AKR49 e AKR120, pode-se considerar que o gene da AKR possui pelo menos nove sequências.

O desenho dos iniciadores para produção dos sgRNAs *in vitro* foi realizado adicionando-se ao *protospacer* as seguintes sequências: ancoragem da T7 RNA polimerase, promotor da T7 RNA polimerase e a parte inicial do *scaffold* (Tabela 3). Note que a sequência PAM não é incluída no iniciador. A Tabela 4 mostra os *donor DNAs* que serão utilizados juntamente com os sgRNAs, sendo específicos para cada um dos sgRNAs escolhidos. Através dos *donor DNAs* serão inseridos códons de parada em diferentes fases de leitura e um sítio de restrição para a enzima *XhoI* por recombinação homóloga.

3.4. ALINHAMENTO GLOBAL

As Figuras S6 a S9 mostram o alinhamento global das sequências de NTR-1, AKR e ADH das cepas CL Brener-Esmeraldo-like e CL Brener Non-Esmeraldo-like de *T. cruzi* feito por meio do alinhamento de sequências par a par usando o Needle (EMBOSS). Podemos observar que as sequências possuem diferenças pontuais de nucleotídeos. Estão destacados de cores distintas as regiões conservadas onde os sgRNAs escolhidos irão se ligar.

Tabela 1 - Número de cópias encontradas na tabela de ortologia e sintenia para a NTR-1, a AKR e a ADH em *T. cruzi* CL Brener. Fonte – TriTrypDB - disponível em <https://tritrypdb.org/tritrypdb/>.

Produto	Gene ID	Organismo	Total de sequências
Nitroreductase	TcCLB.506791.70	<i>T. cruzi</i> CL Brener Non-Esmeraldo-like	2
	TcCLB.510611.60	<i>T. cruzi</i> CL Brener Esmeraldo-like	
Aldo-keto reductase	TcCLB.511287.49	<i>T. cruzi</i> CL Brener Non-Esmeraldo-like	2
	TcCLB.511627.120	<i>T. cruzi</i> CL Brener Esmeraldo-like	
Álcool desidrogenase	TcCLB.511277.60	<i>T. cruzi</i> CL Brener Non-Esmeraldo-like	2
	TcCLB.506357.50	<i>T. cruzi</i> CL Brener Esmeraldo-like	

Fonte - Produzido pelas autoras.

Tabela 2 - sgRNAs selecionados para realizar a deleção dos genes NTR-1, AKR e ADH em *T. cruzi* CL Brener.

sgRNA ID	Sequência do sgRNA (3 últimos nucleotídeos - PAM "NGG" para SpCas9)	Escore total	On-targets (perfect-match non-perfect-but-PAM-match)	Off-targets (perfect-match nonperfect-match)
NTR1_585_revcom	GTAGTACGCATGTCGCAGATAGG	0.85	2 0	0 0
NTR1_375_revcom	CCACGGCTGTAAGTTCAGGGCGG	0.84	2 0	0 0
NTR1_565_revcom	AGGCTGCGCCATAAAGTGAGTGG	0.82	2 0	0 0
AKR_53	AGCTTGGCTTGGGTGTTTGGCGG	0.82	5 0	0 0
AKR49_252	AGGATACGAAAAGACGCTTGCGG	0.79	5 1	0 0
AKR49_229_revcom	TTGTTACCCACACCTCCTCACGG	0.76	5 1	0 0
AKR120_692	CAGCTGCTTACAGTCCTCTGGGG	0.82	4 0	0 0
AKR120_447	AGACGCCAACGGACTTCCTGCGG	0.81	4 0	0 0
AKR120_463_revcom	GAAGTCCGTTGGCGTCTTTGGGG	0.79	4 0	0 0
ADH_349	GCACTTGTC AAGGCTAATGGGGG	0.86	2 0	0 0
ADH_456	AAGCGAAATAACGCGCTTTGCGG	0.83	2 0	0 0
ADH_204	AATGGACCCTGTCGTATTCAAGG	0.82	2 0	0 0

Fonte – Produzido pela autora.

Tabela 3 - Iniciadores para síntese dos sgRNAs para realizar a deleção dos genes NTR-1, AKR e ADH em *T. cruzi* CL Brener usando o kit de transcrição *in vitro* TranscriptAid T7 High Yield Transcription Kit (Thermo Fischer Scientific).

sgRNA ID	Iniciadores 5' → 3' em sequência
	(Ancoragem da T7 RNA polimerase - 8 nucleotídeos, Promotor da T7RNAP - 19 nucleotídeos, <i>protospacer</i> - 20 nucleotídeos, <u>Scaffold</u> – 23 nucleotídeos)
NTR1_585_revcom_F W	GAGAATTGTAATACGACTCACTATAGGGTAGTACGCATGTTCGCAGATGTTTTAGAGCTAGAAATAGC <u>AAG</u>
NTR1_375_revcom_F W	GAGAATTGTAATACGACTCACTATAGGCCACGGCTGTAAGTTCAGGGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
NTR1_565_revcom_F W	GAGAATTGTAATACGACTCACTATAGGAGGCTGCGCCATAAAGTGAGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
ARK49_53	GAGAATTGTAATACGACTCACTATAGGAGCTTGGCTTGGGTGTTTGGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
AKR49_252	GAGAATTGTAATACGACTCACTATAGGAGGATACGAAAAGACGCTTGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
AKR49_229_revcom	GAGAATTGTAATACGACTCACTATAGGTTGTTACCCACACCTCCTCAGTTTTAGAGCTAGAAATAGC <u>AAG</u>
AKR120_692	GAGAATTGTAATACGACTCACTATAGGCAGCTGCTTACAGTCCTCTGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
AKR120_447	GAGAATTGTAATACGACTCACTATAGGAGACGCCAACGGACTTCTGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
AKR120_463_revcom	GAGAATTGTAATACGACTCACTATAGGGAAGTCCGTTGGCGTCTTTGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
ADH_349	GAGAATTGTAATACGACTCACTATAGGGCACTTGTCAAGGCTAATGGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
ADH_456	GAGAATTGTAATACGACTCACTATAGGAAGCGAAATAACGCGCTTTGGTTTTAGAGCTAGAAATAGC <u>AAG</u>
ADH_204	GAGAATTGTAATACGACTCACTATAGGAATGGACCCTGTCGTATTTCAGTTTTAGAGCTAGAAATAGC <u>AAG</u>
Scaffold_RV	<u>ggatccAAAAAAGCACCGACTCGGTGCCAC</u>

Fonte – Produzido pelas autoras.

Tabela 4 - Donor DNAs para inserção dos códons de parada nas sequências codificantes dos genes NTR-1, AKR e ADH de *T. cruzi* CL Brener.

sgRNA ID	Donor DNAs para inserção dos códons de parada (UTR, códons de parada – em negrito, sítio de XhoI – em minúsculo, UTR)
NTR-1_585_revcom	+strand:TACCACTCACTTTATGGCGCAGCCTATCTGTAGATAGATAGctcgagCGAC ATGCGTACTACCTACTTCACGGTGGC
NTR-1_375_revcom	+strand:GCCATGACGGTGCGGGCTCCTACCGCCCTGTAGATAGATAGctcgagAAC TTACAGCCGTGGGTGGCTGTTGTCATA
NTR-1_565_revcom	+strand:ATGGGGCTGGAGAGTGGGTACTACCACTCATAGATAGATAGctcgagCTTT ATGGCGCAGCCTATCTGCGACATGCG
ADH_349	- strand:CAGACCCTCGTAGTCATAGATGGTTCCTCCctcgag CTATCTATCTA ATTAGCCTTGAC
ADH_456	-strand:GACGTGTCGCTTCTCGTCCGTAATAACCGCctcgag CTATCTATCTAAA AG CGCGTTATTTGCTTGCTGTGCCCGA
ADH_204	-strand:GTGGGCAGTTGGGTTCCGGATGCACATCCTTctcgag CTATCTATCTAGA AT ACGACAGGGTCCATTTGACACTTTTT
AKR49_60	-strand:GACCGCATTGCGCGTCTCGGCTCCGTCTGctcgag CTATCTATCTATGC CCGCCAAACACCCAAGCCAAGCTGCGG
AKR49_252	-strand:GAGCAGCTCACGGCTGCGTTCAAAGGCCGctcgag CTATCTATCTAAA AG CGTCTTTTCGTATCCTTGATCTGAATT
AKR49_229_revcom	+strand:GGGATCCGCGAATCGGGAGTGCCCCGTGAGTAGATAGATAGctcgagGA GGTGTGGGTAACAACAAAAGTATGGAAT
AKR120_692	-strand:CGGGTGTGTAATCTTGAAGGCATCCCCAGctcgag CTATCTATCTAAG G ACTGTAAGCAGCTGTCACTATTTGATT
AKR120_447	-strand:ATCGATAAGTTTGACACTATCATCAACCGCctcgag CTATCTATCTAAG GA AGTCCGTTGGCGTCTTTGGGGTATTT
AKR120_463_revcom	+strand:TTCAAGACCGATGACGAGAAATACCCCAAATAGATAGATAGctcgagGAC GCCAACGGACTTCCTGCGGTTGATGAT

Fonte – Produzido pela autora.

4 . DISCUSSÃO

O *T. cruzi* possui uma grande diversidade genética e, por isso, é classificado em seis *Discrete Typing Units* (DTUs): *T. cruzi* I é associado com o ciclo de transmissão silvestre e o *T. cruzi* II, que possui cinco subgrupos (TcIV, TcII, TcIII, TcV e TcVI), é associado com o ciclo de transmissão doméstica (BRISSE; BARNABE;

TIBAYRENCB, 2000; LEWIS *et al.*, 2011; ZINGALES *et al.*, 2009; ZINGALES *et al.*, 2012). Existe uma sétima DTU, chamada de TcBat, sendo um genótipo de parasitos infectivos de morcegos (HAMILTON *et al.*, 2011).

A cepa CL Brener é um membro do subgrupo TcVI, mas é considerada híbrida de dois genótipos (ATWOOD *et al.*, 2005), entre os subgrupos TcII e TcIII, sendo este último considerado um híbrido derivado do *T. cruzi* I

(WESTENBERGER *et al.*, 2005; ZINGALES *et al.*, 2009). Esse parasito possui pares de genes alelos distintos em vários *locis*, com pares de cromossomos homólogos de tamanhos diferentes (STURM *et al.*, 2003; PEDROSO; CUPOLILLO; ZINGALES, 2003).

É estimado que o genoma da cepa CL Brener do *T. cruzi* seja composto por cerca de 22.000 genes que codificam proteínas, sendo que pelo menos 50% do genoma desse parasito é composto de sequências repetitivas e polimórficas (ARNER *et al.*, 2007; EL-SAYED *et al.*, 2005; ZINGALES *et al.*, 2012). Os genes codificadores de proteínas são organizados em longos grupos de genes na mesma fita de DNA (EL-SAYED *et al.*, 2005; ZINGALES *et al.*, 2009). Weatherly, Boehlke e Tarleton (2009) aprimoraram a anotação dos genes de haplótipos previstos da cepa CL Brener (cromossomos Esmeraldo-like e Esmeraldo-non-like), resultando em 41 cromossomos montados.

As funções de várias proteínas do *T. cruzi* ainda não são conhecidas e os principais fatores que contribuem para esse cenário são a grande variação genotípica e fenotípica dentro da espécie; a complexidade do genoma, com um grande número de famílias multigênicas e genes com múltiplas cópias; e ausência da maquinaria do RNAi nesses parasitos (DAROCHA *et al.*, 2004; BURLE-CALDAS *et al.*, 2015). Estão disponíveis métodos para expressar genes exógenos, superexpressar genes endógenos e deletar genes em *T. cruzi*, mas tais metodologias são laboriosas, demoradas e possuem uma baixa taxa de sucesso em gerar mutantes nesses parasitos (PENG e TARLETON, 2015). Sendo assim, o sistema CRISPR/Cas9 provou ser vantajoso pela sua fácil aplicação e alta eficácia em realizar modificações gênicas nesses parasitos (PENG e TARLETON, 2015; LANDER *et al.*, 2016).

Um estudo realizado por Peng *et al.*, (2014) demonstrou sucesso ao realizar a deleção e a inserção de genes através da recombinação homóloga utilizando o sistema CRISPR/Cas9, tendo alta taxa de

fenótipos mutantes e diminuição da expressão proteica em um curto período de tempo. Nesse trabalho os autores observaram um efeito citotóxico da expressão da Cas9 nos parasitos mutantes, no entanto esse efeito não foi observado em nenhum dos trabalhos posteriores de CRISPR/Cas9 em *T. cruzi*. Em seguida, vários outros autores também realizaram a edição gênica do *T. cruzi* utilizando o CRISPR/Cas9 com sucesso, realizando alterações e aprimoramentos no protocolo (LANDER *et al.*, 2016; MEDEIROS *et al.*, 2017; BURLE-CALDAS *et al.*, 2018).

A consulta na tabela de ortologia e sintenia do TriTrypDB permite que todas as cópias ortólogas e parálogas que estejam anotadas para os genes da NTR-1, ADH e AKR sejam devidamente identificadas (ASLETT *et al.*, 2009). Por outro lado, o alinhamento baseado no tBLASTn compara a sequência de aminoácidos (*query*) com um banco de dados de sequências de nucleotídeos (*subject*) traduzidos em todas as fases de leitura. O algoritmo BLAST funciona através do alinhamento de sequências par a par e realiza a busca por similaridade local, encontrando possíveis homólogos em um banco de dados de nucleotídeos ou proteínas. Como o BLAST é um algoritmo capaz de realizar buscas baseadas em alinhamentos rapidamente, ele busca alinhamentos estatisticamente significativos que contêm pares de segmentos de alta pontuação (HSP, *high-scoring segment pairs*), pontuação essa que é atribuída para os *matches* e *mismatches* entre o *query* e o *subject*. (ALTSCHUL *et al.*, 1990; VERLI, 2014). Ao realizar o tBLASTn, a sequência de aminoácidos foi comparada com um banco de dados de nucleotídeos, a fim de encontrar sequências de alta similaridade mesmo no caso das sequências apresentarem algumas diferenças, pois sabe-se que nem todas as mutações de nucleotídeos resultam em uma mudança na sequência de aminoácidos. Essa estratégia permite identificar as sequências que codificam a NTR-1, a AKR e a ADH a nível de genoma, mesmo que porventura não estejam

anotadas (CAMIOLO e PORCEDDU, 2018). Neste trabalho alguns parâmetros foram adotados para filtrar os alinhamentos e obter os melhores resultados, como por exemplo determinar um *e-value* de 10^{-4} , determinar que o número máximo de alinhamentos seria igual a 50 e manter o filtro de baixa complexidade desligado. Nos resultados foram observados a identidade e a cobertura, ambos com valores acima de 90% nos dois alinhamentos. Dessa forma, ao comparar os resultados do tBLASTn com as sequências já anotadas foi possível verificar que não havia cópias dos genes NTR-1 e ADH que ainda não estavam anotadas nos genomas de CL Brener publicados até a presente data. Já no caso da AKR foram retornadas seis sequências através do tBLASTn, resultado que pode indicar tanto erros nas anotações quanto erros na montagem do genoma após o sequenciamento.

No presente trabalho propomos utilizar o método de edição CRISPR/Cas9 descrito por Burle-Caldas *et al.*, (2018), de acordo com o qual foram desenhados sgRNAs e *donor DNAs* para deleção da NTR-1, da ADH e da AKR. A deleção da PGFS, outra enzima relacionada com o fenótipo de resistência, já foi realizada por nosso grupo utilizando essa mesma metodologia (SANTI *et al.*, in prep). No protocolo descrito por de Burle-Caldas *et al.*, (2018) são gerados clones de epimastigotas da cepa CL Brener expressando constitutivamente a nuclease SpCas9 em fusão com a proteína verde fluorescente intensificada (eGFP). O plasmídeo pROCK contendo a sequência codificante da SpCas9 integra no genoma do parasito no *locus* da β -tubulina (DAROCHA *et al.*, 2004b). As células resistentes ao antibiótico G418 e que expressam a SpCas9 e a fluorescência do eGFP foram previamente selecionadas (BURLE-CALDAS *et al.*, 2018) e serão posteriormente transfectadas com os sgRNAs transcritos *in vitro* que reconhecem os genes alvo e também com os *donor DNAs* que possuem os códons de parada e o sítio da enzima *XhoI*. Em nosso trabalho essa enzima foi escolhida por não cortar

nenhuma das sequências da NTR-1, AKR ou ADH (resultados não mostrados), de forma a se poder diferenciar com facilidade as sequências selvagens (WT) das sequências editadas. Uma representação esquemática das transfecções e seus constituintes está exposta na Figura 1.

Para a busca dos possíveis sgRNAs foi utilizado o EuPaGDT, ferramenta que busca todas as sequências PAM (que varia de acordo com o tipo de Cas9 escolhida) tanto na fita senso quanto na fita anti-senso da sequência de interesse fornecida pelo usuário. Além disso, é informado ao programa em qual genoma deseja-se fazer a busca, de forma que, para cada um dos sgRNAs possíveis, são informados todos os *targets* e *off-targets* daquele genoma. Os sgRNAs são pontuados de acordo com sua capacidade de reconhecer os alvos específicos e ao mesmo tempo evitar os *off-targets*. Além disso, a pontuação também é baseada na eficiência do sgRNA (que é calculada considerando o conteúdo GC e a posição dos nucleotídeos) e na possibilidade da quebra naquela região ser reparada por micro-homologia (PENG e TARLETON, 2015).

O *Pairwise Sequence Alignment* foi realizado usando o programa Needle (EMBOSS), que realiza a leitura de duas sequências de aminoácidos ou nucleotídeos (previamente selecionadas pelo usuário) e retorna o alinhamento global ideal das mesmas. Esse programa utiliza o algoritmo Needleman-Wunsch, que tem a capacidade de calcular o melhor escore e alinhamento para duas sequências através da dinâmica de pontuações ao encontrar alinhamentos perfeitos, alinhamento de grande similaridade e alinhamento de baixa similaridade (NEEDLEMAN e WUNSCH, 1970; KRUSKAL, 1983). Esse programa pode ser comparado com o CLUSTAL Omega, que utiliza de um método heurístico para realizar alinhamentos múltiplos de sequências levando em consideração a relação evolutiva entre essas sequências. É realizado o alinhamento par a par (considerando duas sequências

de cada vez) de todos os possíveis pares, verificando os números de caracteres diferentes entre essas sequências, construindo assim a filogenia (alinhamento) (VERLI, 2014). No presente trabalho, o alinhamento global feito por meio do Needle (EMBOSS) foi utilizado para encontrar regiões conservadas nas sequências dos genes da NTR-1, da AKR e da ADH a fim de verificar se os sgRNAs escolhidos reconhecem todas as sequências alvo com sucesso. Caso a mesma análise precisasse ser realizada para mais de duas sequências poderia ser utilizado, por exemplo, o programa CLUSTAL Omega (THOMPSON; HIGGINS; GIBSON, 1994; VERLI, 2014).

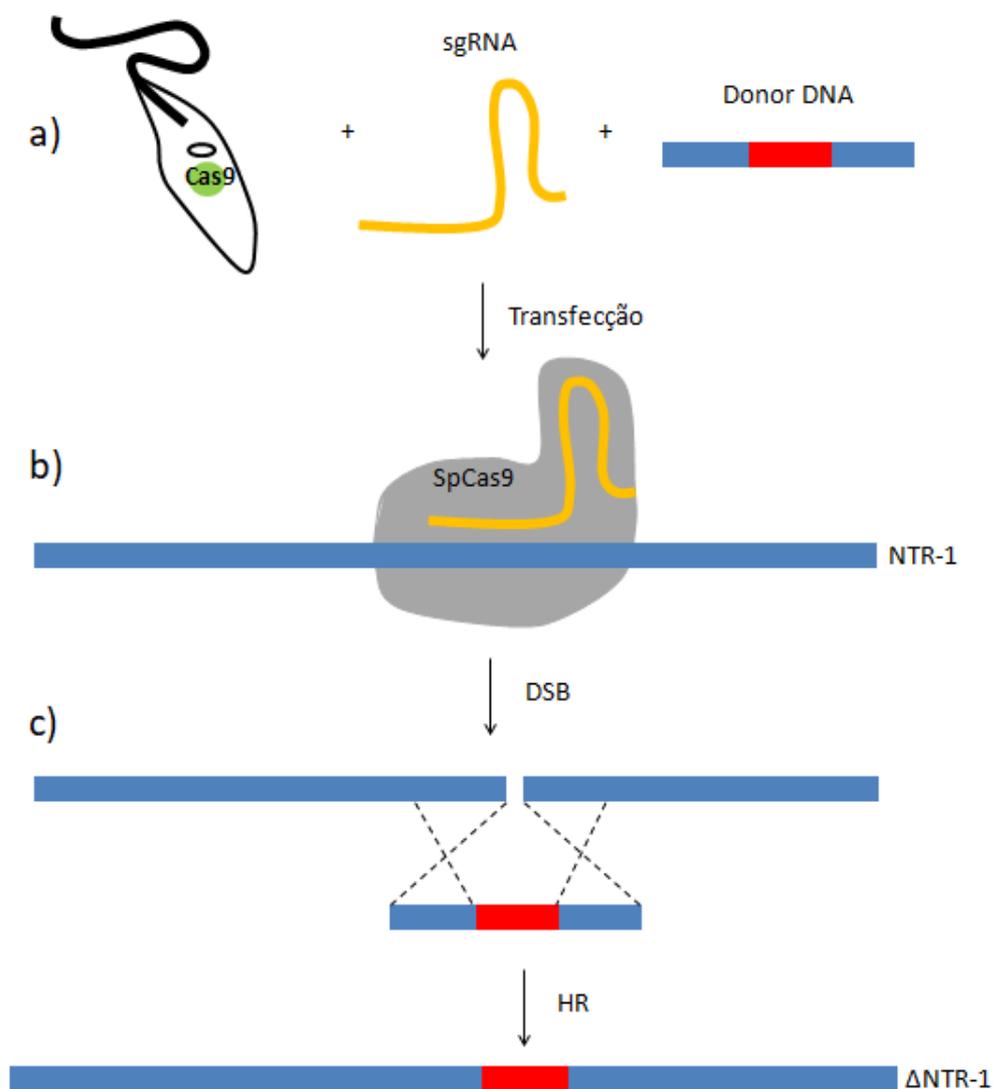
O alinhamento das sequências que codificam a NTR-1 (Fig. S6) e a ADH (Fig.S7) entre as cepas de *T. cruzi* CL Brener Esmeraldo-like e CL Brener Non-Esmeraldo-like demonstrou várias regiões conservadas entre si e poucas variações entre os nucleotídeos, indicando que os sgRNAs escolhidos serão capazes de reconhecer ambas sequências. No caso da AKR foram realizados dois alinhamentos diferentes para demonstrar as regiões de reconhecimento dos sgRNAs, já que eles não são os mesmos para todas as sequências (Fig. S8 e Fig. S9).

Semelhante aos resultados do tBLASTn, na busca pelos sgRNAs foram identificadas duas sequências de NTR-1 e duas de ADH a serem reconhecidas por seus respectivos sgRNAs. Por outro lado, a busca pelos sgRNAs para deleção da AKR revelou que pelo menos nove sequências deverão ser reconhecidas pelos sgRNAs, sendo cinco ou seis sequências AKR49 e quatro sequências AKR120. Esses resultados, em conjunto com os resultados obtidos pelo tBLASTn, indicam que essa enzima possui múltiplas cópias. Dessa forma, a fim de desenhar sgRNAs para AKR de

forma adequada, análises adicionais devem ser realizadas, como por exemplo a análise do sequenciamento do genoma da CL Brener por PACBIO. Nessa estratégia de sequenciamento são obtidas leituras longas que favorecem a correta montagem do genoma, evitando a sobreposição de sequências e a ocorrência de sequências não montadas. Esse tipo de abordagem é a ideal quando se precisa saber com exatidão o número de cópias de um gene, informação essencial para realizar a deleção com sucesso (EID *et al.*, 2009).

A estratégia de deleção utilizando o CRISPR-Cas9 apresentada neste trabalho mostrou-se altamente eficiente em outros estudos (PENG *et al.*, 2015; LANDER *et al.*, 2015; MEDEIROS *et al.*, 2017; BURLE-CALDAS *et al.*, 2018). O desenho dos sgRNAs e dos *donor DNAs* foram feitos seguindo o protocolo de Burle-Caldas *et al.* (2018) e utilizando a mesma ferramenta (EuPaGDT) para busca dos sgRNAs (PENG e TARLETON, 2015). Devido às características dos sgRNAs já apresentadas e da eficiência do protocolo, é esperado que seja possível realizar a deleção dos genes NTR-1 e ADH na cepa CL Brener de *T. cruzi* utilizando essa metodologia, caso os genes não sejam essenciais para os parasitos. Sabe-se que as tentativas de deletar genes essenciais falham, pois cópias desses genes são mantidas no genoma dos parasitos via amplificação ou aneuploidia (CRUZ; TITUS; BEVERLEY, 1993). Em caso de sucesso na deleção, os parasitos editados poderão ser utilizados para determinar a função dessas enzimas no parasito e sua importância nas vias relacionadas à resistência aos fármacos.

Figura 1 – Representação esquemática da deleção da NTR-1 na cepa CL Brener do *T. cruzi*.



a) componentes utilizados na transfecção: parasito expressando a Cas9, sgRNA transcrito *in vitro* e *donor DNA*; b) representação do pareamento do sgRNA com o sítio alvo, seguido pela clivagem da dupla fita de DNA (DSB) pela Cas9; c) recombinação homóloga (HR) a partir do *template* fornecido pelo *donor DNA*, gerando parasitos nocautes para NTR-1 (Δ NTR-1). A mesma estratégia será adotada para deleção da AKR e da ADH.

Fonte: Produzido pelas autoras.

REFERÊNCIAS

- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, 1990.
- ANDRADE, H. M. et al. Proteomic analysis of trypanosoma cruzi resistance to benznidazole. **Journal of Proteome Research**, v. 7, n. 6, p. 2357–2367, 2008.
- ARGOLO, A. M. et al. **Doença de Chagas e seus principais vetores no Brasil**. [s.l.: s.n.].
- ARNER, E. et al. Database of Trypanosoma cruzi repeated genes: 20 000 additional gene variants. **BMC Genomics**, v. 8, p. 1–15, 2007.
- ASH, C.; JASNY, B. R. Trypanosomatid genomes. **Science**, v. 309, n. 5733, p. 399–400, 2005.
- ASLETT, M. et al. TriTrypDB: A functional genomic resource for the Trypanosomatidae. **Nucleic Acids Research**, v. 38, n. SUPPL.1, p. 457–462, 2009.
- ATWOOD, J. A. et al. Microbiology: The Trypanosoma cruzi proteome. **Science**, v. 309, n. 5733, p. 473–476, 2005.
- BARRANGOU, R.; DOUDNA, J. A. Applications of CRISPR technologies in research and beyond. **Nature Biotechnology**, v. 34, n. 9, p. 933–941, 2016.
- BRISSE, S.; BARNABÉ, C.; TIBAYRENC, M. Identification of six Trypanosoma cruzi phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. **International Journal for Parasitology**, v. 30, n. 1, p. 35–44, 2000.
- BOLOTIN, A. et al. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. **Microbiology**, v. 151, n. 8, p. 2551–2561, 2005.
- BURLE-CALDAS, G. A. et al. Assessment of two CRISPR-Cas9 genome editing protocols for rapid generation of Trypanosoma cruzi gene knockout mutants. **International Journal for Parasitology**, v. 48, n. 8, p. 591–596, 2018.
- CAI, Y. et al. CRISPR/Cas9-mediated deletion of large genomic fragments in Soybean. **International Journal of Molecular Sciences**, v. 19, n. 12, 2018.
- CAMIOLO, S.; PORCEDDU, A. **Brachypodium Genomics**. New York, NY: Springer New York, 2018. v. 1667
- CAMPOS, F. M. F. et al. Characterization of a gene encoding alcohol dehydrogenase in benznidazole-susceptible and -resistant populations of Trypanosoma cruzi. **Acta Tropica**, v. 111, n. 1, p. 56–63, 2009.
- CAMPOS, M. C. O. et al. Benznidazole-resistance in Trypanosoma cruzi: Evidence that distinct mechanisms can act in concert. **Molecular and Biochemical Parasitology**, v. 193, n. 1, p. 17–19, 2014.
- CHAGAS, C. “Nova Especie Morbida Do Homem Produzida Por Um Trypanozoma (Trypanozoma Cruzi).” **Brazil-Medico**, no. 16, 1909.
- COURA, J. R. Tripanosomose, doença de chagas. **Ciênc. cult. (São Paulo)**, v. 55, n. 1, p. 30–33, 2003.
- CRUZ, A. K.; TITUST, R.; BEVERLEY, S. M. Plasticity in chromosome number and testing of essential genes in Leishmania by targeting (tetraploid/population biology/aneuploidy/dihydrofolate reductase-thymidylate synthase/protozoan parasite). **Proc. Nati. Acad. Sci. USA**, v. 90, n. February, p. 1599–1603, 1993.
- DAROCHA, W. D. et al. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in Trypanosoma cruzi. **Molecular and Biochemical Parasitology**, v. 133, n. 2, p. 175–186, 2004.
- DAROCHA, W. D. et al. Expression of exogenous genes in Trypanosoma cruzi: Improving vectors and electroporation protocols. **Parasitology Research**, v. 92, n. 2, p. 113–120, 2004.

- DELTCHEVA, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. **Nature**, v. 471, n. 7340, p. 602–607, 2011.
- EID, J. et al. Real-time DNA sequencing from single polymerase molecules. **Science**, v. 323, n. 5910, p. 133–138, 2009.
- FILARDI, L. S.; BRENER, Z. Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas disease. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 81, n. 5, p. 755–759, 1987.
- GARAVAGLIA, P. A. et al. Identification, cloning and characterization of an aldo-keto reductase from *Trypanosoma cruzi* with quinone oxido-reductase activity. **Molecular and Biochemical Parasitology**, v. 173, n. 2, p. 132–141, 2010.
- GARCIA-SALCEDO, J. A. et al. New approaches to overcome transport related drug resistance in trypanosomatid parasites. **Frontiers in Pharmacology**, v. 7, n. SEP, p. 1–14, 2016.
- GARNEAU, J. E. et al. The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. **Nature**, v. 468, n. 7320, p. 67–71, 2010.
- GASIUNAS, G. et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. **Proceedings of the National Academy of Sciences of the United States of America**, v. 109, n. 39, p. 2579–2586, 2012.
- GONZÁLEZ, L. et al. Aldo-keto reductase and alcohol dehydrogenase contribute to benznidazole natural resistance in *Trypanosoma cruzi*. **Molecular Microbiology**, v. 106, n. 5, p. 704–718, 2017.
- GROOM, Z.; PROTOPAPAS, A. D.; ZOCHIOS, V. Tropical diseases of the myocardium: A review. **International Journal of General Medicine**, v. 10, p. 101–111, 2017.
- HAMILTON, P. B. et al. Identification and lineage genotyping of South American trypanosomes using fluorescent fragment length barcoding. **Infection, Genetics and Evolution**, v. 11, n. 1, p. 44–51, 2011.
- HORVATH, P.; BARRANGOU, R. CRISPR/Cas, the immune system of Bacteria and Archaea. **Science**, v. 327, n. 5962, p. 167–170, 2010.
- ISHINO, Y. et al. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. **Journal of Bacteriology**, v. 169, n. 12, p. 5429–5433, 1987.
- JANSEN, R. et al. Identification of genes that are associated with DNA repeats in prokaryotes. **Molecular Microbiology**, v. 43, n. 6, p. 1565–1575, 2002.
- JINEK, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. **Science**, v. 337, n. 6096, p. 816–821, 2012.
- KRUSKAL, J. B. An Overview of Sequence Comparison: Time Warps, String Edits, and Macromolecules. **SIAM Review**, v. 25, n. 2, p. 201–237, 1983.
- LANDER, N.; CHIURILLO, M. A.; DOCAMPO, R. Genome Editing by CRISPR/Cas9: A Game Change in the Genetic Manipulation of Protists. **Journal of Eukaryotic Microbiology**, v. 63, n. 5, p. 679–690, set. 2016.
- LANDER, N. et al. CRISPR/Cas9-induced disruption of paraflagellar rod protein 1 and 2 genes in *Trypanosoma cruzi* reveals their role in flagellar attachment. **mBio**, v. 6, n. 4, p. 1–12, 2015.
- LANDER, N. et al. CRISPR/Cas9-mediated endogenous C-terminal tagging of *Trypanosoma cruzi* genes reveals the acidocalcisome localization of the inositol 1,4,5-trisphosphate receptor. **Journal of Biological Chemistry**, v. 291, n. 49, p. 25505–25515, 2016.

- LEWIS, M. D. et al. Recent, independent and anthropogenic origins of *Trypanosoma cruzi* hybrids. **PLoS Neglected Tropical Diseases**, v. 5, n. 10, 2011.
- LOURDES, R. DE A. E. **Aspectos parasitológicos, imunológicos e moleculares da resposta dependente do receptor do tipo Toll 9 na infecção experimental por cepas de diferentes linhagens de *Trypanosoma cruzi***. [s.l.] Universidade Federal de Minas Gerais, 2013.
- MCCALLA, D. R.; REUVERS, A.; KAISER, C. Breakage of Bacterial DNA by Nitrofurans Derivatives. **Cancer Research**, v. 31, n. 12, p. 2184–2188, 1971.
- MEDEIROS, L. C. S. et al. Rapid, selection-free, high-efficiency genome editing in protozoan parasites using CRISPR-cas9 ribonucleoproteins. **mBio**, v. 8, n. 6, p. 1–15, 2017.
- MEJIA, A. M. et al. Benznidazole-resistance in *trypanosoma cruzi* is a readily acquired trait that can arise independently in a single population. **Journal of Infectious Diseases**, v. 206, n. 2, p. 220–228, 2012.
- MURTA, S. M. F. et al. Deletion of copies of the gene encoding old yellow enzyme (TcOYE), a NAD(P)H flavin oxidoreductase, associates with in vitro-induced benznidazole resistance in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 146, n. 2, p. 151–162, 2006.
- NAEEM, M. et al. Latest Developed Strategies to Minimize the Off-Target Effects in CRISPR-Cas-Mediated Genome Editing. **Cells**, v. 9, n. 7, p. 1–23, 2020.
- NEEDLEMAN, S. B.; WUNSCH, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. **Journal of Molecular Biology**, v. 48, n. 3, p. 443–453, 1970.
- HSU, P. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. **Nature Biotechnology**, v. 31, n. 9, p. 827–832, 2013.
- PEDROSO, A.; CUPOLILLO, E.; ZINGALES, B. Evaluation of *Trypanosoma cruzi* hybrid stocks based on chromosomal size variation. **Molecular and Biochemical Parasitology**, v. 129, n. 1, p. 79–90, 2003.
- PENG, D. et al. CRISPR-Cas9-mediated single-gene and gene family disruption in *Trypanosoma cruzi*. **mBio**, v. 6, n. 1, p. 1–11, 2015
- PENG, D.; TARLETON, R. EuPaGDT: a web tool tailored to design CRISPR guide RNAs for eukaryotic pathogens. **Microbial Genomics**, v. 1, n. 4, p. 1–7, 2015.
- PEREIRA, B. I. et al. TRANSFUSÃO DE SANGUE Qual o Risco nos Países Não Endêmicos? **Acta Medica Portuguesa**, n. December, p. 897–906, 2011
- PÉREZ-MOLINA, J. A.; MOLINA, I. Chagas disease. **The Lancet**, v. 391, n. 10115, p. 82–94, 2018.
- PORCEDDU, S. C. AND A. Identification of Pseudogenes in *Brachypodium distachyon* Chromosomes. v. 1667, n. page 114, p. 149–171, 2017.
- PETRAVICIUS, P. O. et al. Mapping benznidazole resistance in trypanosomatids and exploring evolutionary histories of nitroreductases and ABCG transporter protein sequences. **Acta Tropica**, v. 200, n. April, p. 105161, 2019.
- REID, M. F.; FEWSON, C. A. Molecular characterization of microbial alcohol dehydrogenases. **Critical Reviews in Microbiology**, v. 20, n. 1, p. 13–56, 1994.
- ROBERTS, A. J. et al. A role for trypanosomatid aldo-keto reductases in methylglyoxal, prostaglandin and isoprostane metabolism. **Biochemical Journal**, v. 475, n. 16, p. 2593–2610, 2018.
- SANDER, J. D.; JOUNG, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. **Nature Biotechnology**, v. 32, n. 4, p. 347–350, 2014.

- SARMENTO, R. R. **Interação do Trypanosoma cruzi com células da resposta imune inata.** [s.l: s.n.].
- SANTI, A. M. M. et al. The role of prostaglandin F(2)alpha synthase in *Trypanosoma cruzi* CL Brener drug resistance. *in prep*
- STREETER, A. J.; HOENER, B. ANN. **Evidence for the Involvement of a Nitrenium Ion in the Covalent Binding of Nitrofurazone to DNA** *Pharmaceutical Research: An Official Journal of the American Association of Pharmaceutical Scientists*, 1988.
- STURM, N. R. et al. Evidence for multiple hybrid groups in *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 33, n. 3, p. 269–279, 2003.
- THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673–4680, 1994.
- TROCHINE, A. et al. Benznidazole Biotransformation and Multiple Targets in *Trypanosoma cruzi* Revealed by Metabolomics. **PLoS Neglected Tropical Diseases**, v. 8, n. 5, 2014.
- URBINA, J. A.; DOCAMPO, R. Specific chemotherapy of Chagas disease: Controversies and advances. **Trends in Parasitology**, v. 19, n. 11, p. 495–501, 2003.
- VERLI, H. **Bioinformática: da Biologia à Flexibilidade Molecular.** [s.l: s.n.]. v. 53
- WEATHERLY, D. B.; BOEHLKE, C.; TARLETON, R. L. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. **BMC Genomics**, v. 10, p. 1–13, 2009.
- WESTENBERGER, S. J. et al. Two hybridization events define the population structure of *Trypanosoma cruzi*. **Genetics**, v. 171, n. 2, p. 527–543, 2005.
- WIEDENHEFT, B.; STERNBERG, S. H.; DOUDNA, J. A. RNA-guided genetic silencing systems in bacteria and archaea. **Nature**, v. 482, n. 7385, p. 331–338, 2012.
- WILKINSON, S. R. et al. A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, n. 13, p. 5022–5027, 2008.
- WILKINSON, S. R.; KELLY, J. M. Trypanocidal drugs: Mechanisms, resistance and new targets. **Expert Reviews in Molecular Medicine**, v. 11, n. October 2009, 2009.
- ZHANG, X. H. et al. Off-target effects in CRISPR/Cas9-mediated genome engineering. **Molecular Therapy - Nucleic Acids**, v. 4, n. 11, p. e264, 2015.
- ZINGALES, B. et al. A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. **Corrosion and Protection**, v. 30, n. 6, p. 432–436, 2009
- ZINGALES, B. et al. The revised *Trypanosoma cruzi* subspecific nomenclature: Rationale, epidemiological relevance and research applications. **Infection, Genetics and Evolution**, v. 12, n. 2, p. 240–253, 2012.

ADENDO

Figura S1 – Alinhamento local pelo tBLASTn da enzima NTR-1 nas cepas CL Brener, CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like.

TBLASTN 2.2.28+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Database:
TcruziCLBrenerEsmeraldo-likeGenome;
TcruziCLBrenerNon-Esmeraldo-likeGenome
82 sequences; 65,058,142 total letters

Query= TcCLB.506791.70

Length=313

	Score	E
TcChr28-P organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-...	649	0.0
TcChr28-S organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like...	567	0.0

a)

[Link to Genome Browser](#), Strand = >

TcChr28-P
| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like
| version=2015-12-07 | length=853233 | SO=chromosome
Length=853233

Score = 649 bits (1673), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 313/313 (100%), Positives = 313/313 (100%), Gaps = 0/313 (0%)
Frame = +2

Query	1	MKRNGIKRGLWDSLILYWRWNRENLLRNVTFAENGRHVIGMDGPVEAGSEKDVGRGNSY	60
		MKRNGIKRGLWDSLILYWRWNRENLLRNVTFAENGRHVIGMDGPVEAGSEKDVGRGNSY	
Sbjct	93185	MKRNGIKRGLWDSLILYWRWNRENLLRNVTFAENGRHVIGMDGPVEAGSEKDVGRGNSY	93364
Query	61	MPPIFSSMPPSPSSSSLPLDTMKRVVHERRSCKRFDPTKPIDLVVSDLLAMTVRAPTAL	120
		MPPIFSSMPPSPSSSSLPLDTMKRVVHERRSCKRFDPTKPIDLVVSDLLAMTVRAPTAL	
Sbjct	93365	MPPIFSSMPPSPSSSSLPLDTMKRVVHERRSCKRFDPTKPIDLVVSDLLAMTVRAPTAL	93544
Query	121	NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVTVVFAQDMEPESNAPAALEMGLSEGY	180
		NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVTVVFAQDMEPESNAPAALEMGLSEGY	
Sbjct	93545	NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVTVVFAQDMEPESNAPAALEMGLSEGY	93724
Query	181	YHSLYGAAYLRHAYYLLHGGPCEAMSHVKAIVSAWYSESTGNAMLSVPRNKQAYAWKQVM	240
		YHSLYGAAYLRHAYYLLHGGPCEAMSHVKAIVSAWYSESTGNAMLSVPRNKQAYAWKQVM	
Sbjct	93725	YHSLYGAAYLRHAYYLLHGGPCEAMSHVKAIVSAWYSESTGNAMLSVPRNKQAYAWKQVM	93904
Query	241	IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRFTVPVIIISVGHGAKNGFHSVRSP	300
		IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRFTVPVIIISVGHGAKNGFHSVRSP	
Sbjct	93905	IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRFTVPVIIISVGHGAKNGFHSVRSP	94084
Query	301	RFPTKHLVRWGKF	313
		RFPTKHLVRWGKF	
Sbjct	94085	RFPTKHLVRWGKF	94123

b)

[Link to Genome Browser](#), Strand = >

TcChr28-S

| organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like

| version=2015-12-07 | length=853233 | SO=chromosome

Length=853233

Score = 567 bits (1462), Expect = 0.0, Method: Compositional matrix adjust.

Identities = 294/313 (94%), Positives = 296/313 (95%), Gaps = 1/313 (0%)

Frame = +1

```

Query 1      MKRNGIKRGLWDSLILYWRWNRENLLRNSTFAENGRHVIGMDGPVEAGSEKDVGRGNSY 60
              M+RN IKR L DSLI YWRWNREN L RN S F ENGRHVIGMDGPVEAGSEK+GRGNS
Sbjct 94180   MRRNDIKRRLDLSISYWRWNRENLSRNFSAFVENGRHVIGMDGPVEAGSEKDMGRGNS 94359

Query 61     MPPIFSSMPPSPSSSSSLP LDTMKRVVHERRSCKRFDPTKPIDLDVVS DLLAMTVRAPTAL 120
              MPP FSSMPPS SSS LD MKRVVHERRSCKRFDPTK IDLDVVS DLLAMTVRAPTAL
Sbjct 94360   MPPFFSSMPPSSSSSLP-LDAMKRVVHERRSCKRFDPTKSIDLDVVS DLLAMTVRAPTAL 94536

Query 121    NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVT VVFAGDMEPESNAPAALEMGLESGY 180
              NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVT VVFAGDMEPESNAPAALEMGLESGY
Sbjct 94537   NLQPWVAVVIHEEEQRETLSRAALGQPQPRDAPVT VVFAGDMEPESNAPAALEMGLESGY 94716

Query 181    YHSLYGAAYL RHAYLLHGGPCEAMSHVKAIVSAWYSESTGNAMLSVPRNKQAYANKQVM 240
              YHSLYGAAYL RHAYLLHGGPCE MSHVKAIVSAWYSESTGNAMLSVPRNKQAYANKQVM
Sbjct 94717   YHSLYGAAYL RHAYLLHGGPCEVMSHVKAIVSAWYSESTGNAMLSVPRNKQAYANKQVM 94896

Query 241    IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRF T V PVIISVGHGAKNGFHSVRSP 300
              IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRF T V PVIISVGHGAKNGFHSVRSP
Sbjct 94897   IPATTFLYLATAAGFDTAILEGFDEAQVRRVAGLPPRF T V PVIISVGHGAKNGFHSVRSP 95076

Query 301    RFPTKHLVRWGKF 313
              RFPTKHLVRWGKF
Sbjct 95077   RFPTKHLVRWGKF 95115

```

Lambda	K	H	a	alpha
0.319	0.134	0.415	0.792	4.96

Lambda	K	H	a	alpha	sigma
0.267	0.0410	0.140	1.90	42.6	43.6

Effective search space used: 4508906896

Database:

TcruziCLBrenerEsmeraldo-likeGenome

Posted date: Sep 25, 2020 10:18 AM

Number of letters in database: 32,529,070

Number of sequences in database: 41

Database:

TcruziCLBrenerNon-Esmeraldo-likeGenome

Posted date: Sep 24, 2020 12:57 PM

Number of letters in database: 32,529,072

Number of sequences in database: 41

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1

Neighboring words threshold: 13

Window for multiple hits: 40

A busca a nível genômico por cópias da NTR-1 retornou apenas duas sequências: a) alinhamento local da cepa CL Brener Non-Esmeraldo-like; b) alinhamento local da cepa CL Brener Esmeraldo-like. Fonte:

produzido pela autora utilizando ferramentas disponíveis no programa TriTrypDB - disponível em <https://tritrypdb.org/tritrypdb/>.

Figura S2 – Alinhamento local pelo tBLASTn da enzima ADH nas cepas CL Brener, CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like.

TBLASTN 2.2.28+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Database:
TcruziCLBrenerEsmeraldo-likeGenome;
TcruziCLBrenerNon-Esmeraldo-likeGenome
82 sequences; 65,058,142 total letters

Query= TcCLB.511277.60

Length=392

	Score	E
TcChr40-P organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-...	810	0.0
TcChr40-S organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like...	805	0.0

a)

[Link to Genome Browser](#), Strand = >

TcChr40-P

| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like
| version=2015-12-07 | length=2036760 | SO=chromosome
Length=2036760

Score = 810 bits (2091), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 392/392 (100%), Positives = 392/392 (100%), Gaps = 0/392 (0%)
Frame = +1

Query	1	MFRFSRPSAAASTFFMPKVSYLGVGALDSATSMTKLGFKKPLLVCOTGIEKAGILETVM	60
Sbjct	816721	MFRFSRPSAAASTFFMPKVSYLGVGALDSATSMTKLGFKKPLLVCOTGIEKAGILETVM	816900
Query	61	DVLKKKCCQMDPVVFKDVHPNPTAHNVETGLELLNKNACDSIIAIGGGSPQDCAKGIALVK	120
Sbjct	816901	DVLKKKCCQMDPVVFKDVHPNPTAHNVETGLELLNKNACDSIIAIGGGSPQDCAKGIALVK	817080
Query	121	ANGGTIYDYEGLDKAEKDQYPLVCINTTSGTASEITRFVITDEKRHVKMVITTDVTPL	180
Sbjct	817081	ANGGTIYDYEGLDKAEKDQYPLVCINTTSGTASEITRFVITDEKRHVKMVITTDVTPL	817260
Query	181	VAIDDAQLMMGPPSLTAATGMDALHVAEAYVSVGASETTCALYAMRLIKQHLSNAV	240
Sbjct	817261	VAIDDAQLMMGPPSLTAATGMDALHVAEAYVSVGASETTCALYAMRLIKQHLSNAV	817440
Query	241	KDGKNLVAREGMCYAQHLAGMAFNSAGLGYVHAMAHQLGGFYDLPHGVCNAILLPHVESY	300
Sbjct	817441	KDGKNLVAREGMCYAQHLAGMAFNSAGLGYVHAMAHQLGGFYDLPHGVCNAILLPHVESY	817620
Query	301	NSFVCCARRLGDVAVQLGVSTTGMDKNAAAAAIDAIRALSKEVQIPSGFEQLGMKEKDIP	360
Sbjct	817621	NSFVCCARRLGDVAVQLGVSTTGMDKNAAAAAIDAIRALSKEVQIPSGFEQLGMKEKDIP	817800
Query	361	ALAESALKDVCAATNPRQGSKEDVMKIYRESM	392
Sbjct	817801	ALAESALKDVCAATNPRQGSKEDVMKIYRESM	817896

b)

[Link to Genome Browser](#), Strand = >
 TcChr40-5
 | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like
 | version=2015-12-07 | length=2036759 | SO=chromosome
 Length=2036759

Score = 805 bits (2080), Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 390/392 (99%), Positives = 391/392 (99%), Gaps = 0/392 (0%)
 Frame = +2

Query	1	MFRFSRPSAAASTFFMPKVSYLGVGALDSATS	SKMTKLGFKKPLLVCOTGIEKAGILETVM	60
Sbjct	818456	MFRFSRPSAAASTFFMPKVSYLGVGALDSATS	SKMTKLGFKKPLLVCOTGIEKAGILETVM	818635
Query	61	DVLKCKCQNDPVPVKDVHPNPTAHNVETGLELLNKNACDSIIAIGGGSPQDCAKGI	ALVK	120
Sbjct	818636	DVLKCKCQNDPVPVKDVHPNPTAHNVETGLELLNKNACDSIIAIGGGSPQDCAKGI	ALVK	818815
Query	121	ANGGTIYDYEGLDKAEKDQYPLVCINTTSGTASEITRFAVITDEKRHVKMVITTD	VTPL	180
Sbjct	818816	ANGGTIYDYEGLDKAEKDQYPLVCINTTSGTASEITRFAVITDEKRHVKMVITTD	VTPL	818995
Query	181	VAIDDAQLMMGQPPSLTAATGMDALTHAVEAYVSVGASETTD	DACALYAMRLIKQHLSNAV	240
Sbjct	818996	VAIDDAQLMMGQPPSLTAATGMDALTHAVEAYVSVGASETTD	DACALYAMRLIKQHLSNAV	819175
Query	241	KDGNLVAREGMCYAQHLAGMAFN	SAGLGYVHAMAHQLGGFYDLP	300
Sbjct	819176	KDGNLVAREGMCYAQHLAGMAFN	SAGLGYVHAMAHQLGGFYDLP	819355
Query	301	NSFVCARRLGDVAVQLGVSTTGMDKNAAAAAIDAIRALSKEVQIP	SGFEQLGMKEKDIP	360
Sbjct	819356	NSFVCARRLGDVAVQLGVSTTGMDKNAAAAAIDAIRALSKEVQIP	SGFEQLGMKEKDIP	819535
Query	361	ALAESALKDVCAATNPRQGSKEDVMKIYRESM	392	
Sbjct	819536	ALAESALKDVCAATNPRQGSKEDVMKIYRESM	819631	

Lambda K H a alpha
 0.318 0.132 0.381 0.792 4.96

Gapped
 Lambda K H a alpha sigma
 0.267 0.0410 0.140 1.90 42.6 43.6

Effective search space used: 6178022805

Database:
 TcruziCLBrenerEsmeraldo-likeGenome
 Posted date: Sep 25, 2020 10:18 AM
 Number of letters in database: 32,529,070
 Number of sequences in database: 41

Database:
 TcruziCLBrenerNon-Esmeraldo-likeGenome
 Posted date: Sep 24, 2020 12:57 PM
 Number of letters in database: 32,529,072
 Number of sequences in database: 41

Matrix: BLOSUM62
 Gap Penalties: Existence: 11, Extension: 1
 Neighboring words threshold: 13
 Window for multiple hits: 40

A busca a nível genômico por cópias da ADH retornou apenas duas sequências: a) alinhamento local da cepa CL Brener Non-Esmeraldo-like; b) alinhamento local da cepa CL Brener Esmeraldo-like. Fonte:

produzido pela autora utilizando ferramentas disponíveis no programa TriTrypDB - disponível em <https://tritrypdb.org/tritrypdb/>.

Figura S3 - *Pairwise Sequence Alignment* usando o Needle (EMBOSS) demonstrando as diferenças entre as seqüências de aminoácidos entre a AKR49 e a AKR120.

```

# Aligned_sequences: 2
# 1: TcCLB.511287.49
# 2: TcCLB.511627.120
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 350
# Identity:      98/350 (28.0%)
# Similarity:   156/350 (44.6%)
# Gaps:        79/350 (22.6%)
# Score: 385.0

TcCLB.511287. 1 -----MNCVNYNCVTLHINSARMPOGLGVHRAQDGAETAN      34
                :.:.:. :.:|...:|...|...|...|...|...
TcCLB.511627. 1 MPGANVLSFKPLGGPVLGVSHR-LPLRNGNSIPQCQGFGRMTPVAGA-  48

TcCLB.511287. 35 AVRWAI EAGYRHIDTACIYVNEKGVGGGIRE----SGVPREEVWVTKKW  80
                ||.:||..|:||||..|.:|:..:|:..:   ...|:|:..:|:|:|
TcCLB.511627. 49 AVEYAIHCGFRHIDCAKAYDNQNAIGEALQRVISTGNLKRLEFLTSKLV  98

TcCLB.511287. 81 NSDQ-----GYEKTAAAFERSRELLGLEIDLYLIHMP-----  113
                .:| |       ..:| | | .       |.:|:| | | | | | | |
TcCLB.511627. 99 PTQHPIHVEKACRETAE-----LRVDYLDLYLIHMPVWVNHSPHFK  141

TcCLB.511287. 114 -----GKKKFVDTNKALEKLYEEKKVRAIGVSNFEPHH      146
                ...|.:| | | |:|:|.:.|:..|:|:|:|:|...|
TcCLB.511627. 142 TDDEKYPKDANGLPAVDDSVKLIIDTRAMCELVDRLNLRVSVGLSNCSEKH  191

TcCLB.511287. 147 LTELFSCKI-RPMVNIQVELHPLFQORTLREFCKQHNIATAWSPHSGSD  195
                :.:|:.....: .:| | | | | | | | | | | | | | | | | | | | | |
TcCLB.511627. 192 INEVMDSGSLVAPVWNIQIELHPALVQRDLLNFHGAIQIVTAAVSPHGMPS  241

TcCLB.511287. 196 R-----TGFLKNHMLGEIAKKHNKSPAQWIRWDIQHGIVTIPKSTNKG  239
                |      .|.:..:|:..:|:.....:|:..:|:.....:|:..:| | | | |
TcCLB.511627. 242 RFTHPDYKGLLSDASLQSISEYSGFSVARLLLNMINLDMHNWIVRSTNKE  291

TcCLB.511287. 240 RIQENFINWDFKLTEEEEMRQIIDLNEDKRIGG----HPDNFFPGGEE---  282
                .|:..|:.....:|:..:|:.....:|:..:|:.....:|:..:| | | | |
TcCLB.511627. 292 HIRSNKASLYALSDPVRMILDRFQE--RVGTIRTINPTDFTRGGGFFD      339

```

Alinhamento global das seqüências de TcAKR de CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo like por Needle (EMBOSS) mostrando as diferenças entre as seqüências, que estão destacadas por pontuações onde: um bastão (|) indica alinhamento perfeito; dois pontos (:) indica que esse sítio pertence a um grupo com alta similaridade; um ponto (.) indica que esse sítio pertence a um grupo com baixa similaridade. Fonte: produzido pelas autoras utilizando ferramentas disponíveis no programa Needle (EMBOSS) https://www.ebi.ac.uk/Tools/psa/emboss_needle/.

Figura S4 – Alinhamento local da AKR-49 pelo tBLASTn contra os genomas das cepas CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like.

```
TBLASTN 2.2.28+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.

Database:
TcruziCLBrenerEsmeraldo-likeGenome;
TcruziCLBrenerNon-Esmeraldo-likeGenome
82 sequences; 65,058,142 total letters

Query= TcCLB.511287.49
Length=282

Score      E
TcChr40-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 590 0.0
TcChr40-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 470 5e-150
TcChr23-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 149 5e-39
TcChr26-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 130 1e-32
TcChr23-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 82.4 7e-17
TcChr26-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 60.1 1e-09

Link to Genome Browser, Strand = >
TcChr40-P
| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like
| version=2015-12-07 | length=2036760 | SO=chromosome
Length=2036760

Score = 590 bits (1520), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 282/282 (100%), Positives = 282/282 (100%), Gaps = 0/282 (0%)
Frame = +2

Query 1      MNCNYNCVTLHNSARMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTACIYNNEKGVG 60
Sbjct 486257 MNCNYNCVTLHNSARMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTACIYNNEKGVG 486436

Query 61     QGIRESGVPREEVWTTKVWNSDQGYEKTAAAFERSRELLGLEIDLYLIHWPGKKKFVD 120
Sbjct 486437 QGIRESGVPREEVWTTKVWNSDQGYEKTAAAFERSRELLGLEIDLYLIHWPGKKKFVD 486616

Query 121    TWKALEKLYEEKKVRAIGVSNFEPHHLTELFKSCKIRPMVNQVELHPLFQQRTRLREFCKQ 180
Sbjct 486617 TWKALEKLYEEKKVRAIGVSNFEPHHLTELFKSCKIRPMVNQVELHPLFQQRTRLREFCKQ 486796

Query 181    HNIAITAWSPLGSGDRTGFLKNHVLGEIAKKHNKSPAQVVIRWDIQHGIVTIPKSTNKGR 240
Sbjct 486797 HNIAITAWSPLGSGDRTGFLKNHVLGEIAKKHNKSPAQVVIRWDIQHGIVTIPKSTNKGR 486976

Query 241    IQENFNWDFKLT EEE MRQIDELNEDKRIGGHPDNFFPGGEE 282
Sbjct 486977 IQENFNWDFKLT EEE MRQIDELNEDKRIGGHPDNFFPGGEE 487102
```

[Link to Genome Browser](#), Strand = >

TcChr40-S

| organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like

| version=2015-12-07 | length=2036759 | SO=chromosome

Length=2036759

Score = 470 bits (1209), Expect = 5e-150, Method: Compositional matrix adjust.

Identities = 228/251 (91%), Positives = 233/251 (93%), Gaps = 4/251 (2%)

Frame = +3

```

Query  32      TANAVRWAIEAGYRHIDTACIYNNKEGVGQGIRESGVPREEVWVTTKVWNSDQGYEKTLA  91
      +  +R A      RH      YNNE+GVGQGIRESGVPREEVWVTTKVWNSDQGYEKTLA
Sbjct  486636   SGGQLRQATATLIRH----TFYNNERGVGQGIRESGVPREEVWVTTKVWNSDQGYEKTLA  486803

Query  92      AFERSRELLGLEIDLYLIHWP GK KKFVDTWKALEKLYEEKVRAIGVSNFEPHHLTELF  151
      AFERSRELLGLEIDLYLIHWP GK KKFVDTWKALEKLYEEKVRAIGVSNFEPHHLTELF
Sbjct  486804   AFERSRELLGLEIDLYLIHWP GK KKFVDTWKALEKLYEEKVRAIGVSNFEPHHLTELF  486983

Query  152     KSCKIRPMVNQVELHPLFQQRTRLREFCKQHNI AITAWSPLGSGDRTGFLKNHVLGEIAKK  211
      +SCKIRPMVNQVELHPLFQQRTRLREFCKQHNI AITAWSPLGSGDRTGFLKNHVLGEIAKK
Sbjct  486984   QSCKIRPMVNQVELHPLFQQRTRLREFCKQHNI AITAWSPLGSGDRTGFLKNHVLGEIAKK  487163

Query  212     HNKSPAQVVIRWDIQHGIVTIPKSTNKGRIQENFNWVDFKLTEEMRQIDELNEDKRIGG  271
      HNKSPAQVVIRWDIQHGIVTIPKS NKGRIQENFNWVDFKLTEE+MRQIDELNEDKR GG
Sbjct  487164   HNKSPAQVVIRWDIQHGIVTIPKSANKGRIQENFNWVDFKLTEEDMRQIDELNEDKRFGG  487343

Query  272     HPDNFFPGGEE  282
      HPDNFFP GEE
Sbjct  487344   HPDNFFPDGEE  487376

```

Score = 112 bits (279), Expect = 2e-26, Method: Compositional matrix adjust.

Identities = 52/62 (84%), Positives = 53/62 (85%), Gaps = 0/62 (0%)

Frame = +2

```

Query  1      MNCNYNCVTLHNSARMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTACIYNNKEGVG  60
      MNCNY+CVTLHNS RMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTA      KG G
Sbjct  486530   MNCNYS CVTLHNSVRMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTAYFLQQRKGRG  486709

Query  61      QG  62
      G
Sbjct  486710   AG  486715

```

[Link to Genome Browser](#), Strand = >

TcChr23-S

| organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like
| version=2015-12-07 | length=655477 | SO=chromosome
Length=655477

Score = 149 bits (376), Expect = 5e-39, Method: Compositional matrix adjust.
Identities = 97/315 (31%), Positives = 150/315 (48%), Gaps = 45/315 (14%)
Frame = +2

```

Query 8      VTLHNSARMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTACIYNNEKGVGQG----I 63
             + L N  +PQ G G +R          A AV +AI  G+RHID A  Y+N+  +G+   I
Sbjct 528689 LPLRNGNSIPQCGFGTYRMTPTVAGA-AVEYAIHCGFRHIDCAKAYDNQNAIGEALQRVI 528865

Query 64     RESGVPREEVWVTTKVMNSDQGYEKTAAAFERSRELLGLEIDLYLIHWP----- 113
             + REE+++T+K+W +DQ          A  +  L  ++Y+DLYLIHWP
Sbjct 528866 STGNLKREELFLTSKLNPTDQHPHVEKACRETLAELRVDYLDLYLIHWPVVMNHSPHFK 529045

Query 114    -----GKKKFVDTWKALEKLYEEKKVRAIGVSNFEPHHLTELFKSKI 156
             K +DTW+A+ +L +  VR++G+SN  H+ E+  +
Sbjct 529046 TDDEKYPKDANGLPAVDDSVKLIIDTWRAMCELVDNRNLVRSVGLSNCSEKHINEVMSDGLS 529225

Query 157    -RPMVNQVELHPLFQORTLREFCKQHNIATAWSPLGSGDR-----TGFLKNHVLGEIA 209
             P+VNQ+ELHP  QR L F  + I  A+SPLG  R          G L + L I+
Sbjct 529226 YAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSPLGMPSRFTHPDYKGLLSDASLQSI 529405

Query 210    KKHNKSPAQVVIRWDIQHGIVTIPKSTNKGRIQENFNWVDFKLTEEEMRQIDELNEDKRI 269
             +  S A++++ W++  V I +STNK  I+ N  + L++  +D  E  R+
Sbjct 529406 EYSGFSVARLLLWNLDMMHNVVIVRSTNKEHIRSNKASLYALSDPVRMILDRFQE--RV 529579

Query 270    GG----HPDNFFPGG 280
             G      +P +F  GG
Sbjct 529580 GTIRTINPTDFTRGG 529624

```

[Link to Genome Browser](#), Strand = >

TcChr26-P

| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like

| version=2015-12-07 | length=801422 | SO=chromosome

Length=801422

Score = 130 bits (328), Expect = 1e-32, Method: Compositional matrix adjust.

Identities = 86/286 (30%), Positives = 137/286 (48%), Gaps = 33/286 (12%)

Frame = -3

```

Query 15      RMPQLGLGVHRAQDGAETANAVRWAI EAGYRHIDTACIYNNEKGVGQGIRESGVPREEVW 74
              R+P +G+G +   G E+ AV A+ G+R +DTA Y+NE+ VG I ESGV REE++
Sbjct 588621  RIPHMGVGTYELY-GDESRIAVLAALRLGFRVLDTAAGYHNEEHVGA AIEESGVQREELF 588445

Query 75      VTTKVVNSDQGYEKT L-AAFERSRELLGLE YIDLYLIHWPG-----KKKFVDT 121
              V K+      E+ +   S L + Y D LIHWPG          +
Sbjct 588444  VVKIAPKAMASEELVEKGIRESVRKLR IAYADCVLIHWPGCGGLKPEEAEGHRAARHRC 588265

Query 122     WKALEKLYEEKKVRAIGVSNFEPHHLTE LFKSCK-----IRPMVNQVELHPLFQ 170
              WK ++ L +E +R +GVSNF P H L +          RP++NQ+ELHPL
Sbjct 588264  WKVMQALKKEGIIRHLGVS NFAPRHFP HLVGNA AEEAHLFDAANPCR PVINQIELHPLCV 588085

Query 171     QRTLREFCKQHNI AITAWSPLGSGDR TGF LKNHVLGEI AKK--HNKSPAQVVIRWDIQHG 228
              Q+ +FC++ +I + +SPL G   +++ VL I K   + V++ W + G
Sbjct 588084  QQEACDFCRERDI ILQQYSPLAQG-HAKLVEHPVLCAIVKDL FPGYTVQDVLLMWGLSQG 587908

Query 229     IVTIPKSTNKGRIQENFNVW-----DFKLTEEEMRQIDELNEDKRI 269
              + +S +K + +N+   + L+E + Q+ L E R+
Sbjct 587907  FCVVRSRSKEHLAQNWAAAKAFFDEGLLSESQKEQLKNLRESMRV 587770

```

[Link to Genome Browser](#), Strand = >

TcChr23-P

| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like

| version=2015-12-07 | length=655477 | SO=chromosome

Length=655477

Score = 82.4 bits (202), Expect = 7e-17, Method: Compositional matrix adjust.

Identities = 53/174 (30%), Positives = 80/174 (46%), Gaps = 32/174 (18%)

Frame = +2

```
Query 8      VTLHNSARMPQLGLGVWRAQDGAETANAVRWAIEAGYRHIDTACIYNNKEKGVGQG----I 63
             + L N  +PQ G G +R      A AV +AI G+RHID A Y+N+ +G      I
Sbjct 528644 LPLRNGNSIPQCGFTYRMNPTVAGA-AVEYAIHCGFRHIDCAKAYDNQNAIGDALQRVI 528820
```

```
Query 64      RESGVPREEVWTTKVMNSDQGYEKTAAAFERSRELLGLEIDLYLIHWP----- 113
             + REE+++T+K+W +DQ      A + L ++Y+DLYLIHWP
Sbjct 528821 STGKLRKREELFLTSKLPDQHPHVEKACRETLAELRVDYLDLYLIHWPVWVNHSPHLK 529000
```

```
Query 114     -----GKKKFVDTWKALEKLYEEKKVRAIGVSNFEPHHLTEL 150
             K +DTW+A+ +L + VR +G      + LT+L
Sbjct 529001 TDDEKYPKDANGLPAVDDSVKLIIDTWRAMCELVDRIWVRQLGSRVVRNILTCL 529162
```

Score = 71.2 bits (173), Expect = 2e-13, Method: Compositional matrix adjust.

Identities = 48/156 (31%), Positives = 76/156 (49%), Gaps = 13/156 (8%)

Frame = +1

```
Query 136     AIGVSNFEPHHLTELFKSCKI-RPMVNQVELHPLFQQRTLREFCKQHNIATAWSPLGSG 194
             ++G+SN H+ E+      + P+VNQ+ELHP QR L F + I A+SPLG
Sbjct 529117 SVGLSNCSEKHINEVMSDGSLYAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSPLGMP 529296
```

```
Query 195     DR-----TGFLKNHVLGEIAKKNKSPAQVIRWDIQHGIVTIPKSTNKGRIQENFNW 248
             R      G L + L I++ S A++++ W++ V I +STNK I+ N
Sbjct 529297 SRFTHPNYKGLLYDASLQSISEYSGFSVARLLLNWNLDMHNWVIVRSTNKEHIRSNAKAS 529476
```

```
Query 249     DFKLTEEEMRQIDELNEDKRIGG----HPDNFFPGG 280
             + L++      +D E R+G      +P +F GG
Sbjct 529477 LYALSDPVRMILDRFQE--RVGTIRTINPTDFTRGG 529578
```

[Link to Genome Browser](#), Strand = >

TcChr26-5

| organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like
| version=2015-12-07 | length=801422 | SO=chromosome
Length=801422

Score = 60.1 bits (144), Expect = 1e-09, Method: Compositional matrix adjust.
Identities = 39/150 (26%), Positives = 69/150 (46%), Gaps = 19/150 (13%)
Frame = -3

```

Query 138  GVSNFEPHHLTELKFKSCK-----IRPMVNQVELHPLFQQRTLREFCKQHIAIT 186
           G  NF P H  L  +                RP++NQ+ELHPL  Q+ + +FC++ +I +
Sbjct 588753 GGVNFAPRHFPHLVGNASEEAHLFDAANPCRPVINQIELHPLCVQEQVDFCRERDIIQ 588574

Query 187  AWSPLGSGDRTGFLKNHVLGEIAKK--HNKSPAQVVIRWDIQHGIVTIPKSTNKGRIQEN 244
           +SPL G  +++ VL I K  +  V++ W + G  + +S +K  + +N
Sbjct 588573 QYSPLAQGH-AKLVEHPVLCIAIVKDLFPGYTVQDVLMLWGLSQGFCVWVRSRSKEHLAQN 588397

Query 245  FNVW-----DFKLTEEEMRQIDELNEDKRI 269
           +                + L+E + Q+ L E R+
Sbjct 588396 WAAAKAFFDEGLLSESQKEQLKNLRESMRV 588307

```

Lambda	K	H	a	alpha
0.318	0.136	0.429	0.792	4.96

Gapped Lambda	K	H	a	alpha	sigma
0.267	0.0410	0.140	1.90	42.6	43.6

Effective search space used: 3858598382

Database:

TcruziCLBrenerEsmeraldo-likeGenome
Posted date: Sep 25, 2020 10:18 AM
Number of letters in database: 32,529,070
Number of sequences in database: 41

Database:

TcruziCLBrenerNon-Esmeraldo-likeGenome
Posted date: Sep 24, 2020 12:57 PM
Number of letters in database: 32,529,072
Number of sequences in database: 41

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1
Neighboring words threshold: 13
Window for multiple hits: 40

A busca a nível genômico por cópias da AKR49 retornou seis sequências.

Fonte:

produzido pelas autoras utilizando ferramentas disponíveis no programa TriTrypDB - disponível em <https://tritrypdb.org/tritrypdb/>.

Figura S5 – Alinhamento local da AKR120 pelo tBLASTn contra os genomas das cepas CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like.

```
TBLASTN 2.2.28+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.

Database:
TcruziCLBrenerEsmeraldo-likeGenome;
TcruziCLBrenerNon-Esmeraldo-likeGenome
      82 sequences; 65,058,142 total letters

Query= TcCLB.511627.120
Length=339

                                Score   E
TcChr23-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 708   0.0
TcChr23-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 372   0.0
TcChr40-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 149   2e-38
TcChr26-P | organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-... 130   3e-32
TcChr40-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 117   1e-27
TcChr26-S | organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like... 64.7  6e-11

Link to Genome Browser, Strand = >
TcChr23-S
| organism=Trypanosoma_cruzi_CL_Brener_Esmeraldo-like
| version=2015-12-07 | length=655477 | SO=chromosome
Length=655477

Score = 708 bits (1827), Expect = 0.0, Method: Compositional matrix adjust.
Identities = 339/339 (100%), Positives = 339/339 (100%), Gaps = 0/339 (0%)
Frame = +2

Query 1      MPGANVLSFKPLGGPVLGVSHRLPLRNGNSIPQCGFGTYRMTPTVAGAAVEYAIHCGFRH 60
Sbjct 528623  MPGANVLSFKPLGGPVLGVSHRLPLRNGNSIPQCGFGTYRMTPTVAGAAVEYAIHCGFRH 528802

Query 61     IDCAKAYDNQNAIGEALQRVISTGNLKREELFLTSKLWPTDQHPHVEKACRETLAELRV 120
Sbjct 528803  IDCAKAYDNQNAIGEALQRVISTGNLKREELFLTSKLWPTDQHPHVEKACRETLAELRV 528982

Query 121    DYLDLYLIHWPVWNHSPHFKTDDKEYPKDANGLPAVDDSVKLIIDTWRAMCELVDNRNLVR 180
Sbjct 528983  DYLDLYLIHWPVWNHSPHFKTDDKEYPKDANGLPAVDDSVKLIIDTWRAMCELVDNRNLVR 529162

Query 181    SVGLSNCSEKHINEVMSDGSLYAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSPLGMP 240
Sbjct 529163  SVGLSNCSEKHINEVMSDGSLYAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSPLGMP 529342

Query 241    SRFTHPDYKGLSDASLQSISEYSGFSVARLLNWNLDMHNWVIVRSTNKEHIRSNAKAS 300
Sbjct 529343  SRFTHPDYKGLSDASLQSISEYSGFSVARLLNWNLDMHNWVIVRSTNKEHIRSNAKAS 529522

Query 301    LYALSDPVRMILDRFQERVGTIRTINPTDFTRGGGSFFD 339
Sbjct 529523  LYALSDPVRMILDRFQERVGTIRTINPTDFTRGGGSFFD 529639
```

[Link to Genome Browser](#), Strand = >

TcChr23-P

| organism=Trypanosoma_cruzi_CL_Brener_Non-Esmeraldo-like

| version=2015-12-07 | length=655477 | SO=chromosome

Length=655477

Score = 372 bits (955), Expect(2) = 0.0, Method: Compositional matrix adjust.
Identities = 175/183 (96%), Positives = 177/183 (97%), Gaps = 0/183 (0%)
Frame = +2

```
Query 1      MPGANVLSFKPLGGPVLGVSHRLPLRNGNSIPQCGFGTYRMTPTVAGAAVEYAIHCGFRH 60
Sbjct 528578 MPGANVLSFKPLGGPVLGVSHRLPLRNGNSIPQCGFGTYRMTPTVAGAAVEYAIHCGFRH 528578

Query 61     IDCAKAYDNQNAIGEALQRVISTGNLKRLEELFTSKLWPTDQHPHVEKACRETLAELRV 120
Sbjct 528758 IDCAKAYDNQNAIG+ALQRVISTG LKREELFTSKLWPTDQHPHVEKACRETLAELRV 528937

Query 121    DYLDLYLIHWPVVWNHSPHFKTDDEKYPKDANGLPAVDDSVKLLIDTWRAMCELVDRLV 180
Sbjct 528938 DYLDLYLIHWPVVWNHSPH KTDDEKYPKDANGLPAVDDSVKLLIDTWRAMCELVDR VR 529117

Query 181    SVG 183
Sbjct 529118 +G
QLG 529126
```

Score = 326 bits (835), Expect(2) = 0.0, Method: Compositional matrix adjust.
Identities = 159/163 (98%), Positives = 160/163 (98%), Gaps = 0/163 (0%)
Frame = +1

```
Query 177    NLVRSVGLSNCSEKHINEVMSDGSLYAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSP 236
Sbjct 529105 NLGASVGLSNCSEKHINEVMSDGSLYAPVVNQIELHPALVQRDLLNFHGANQIVTAAYSP 529284

Query 237    LGMPSRFTHPYKGLLDASLQSISEYSGFSVARLLLNWNLDMHNWVIVRSTNKEHIRSN 296
Sbjct 529285 LGMPSRFTHPYKGLLDASLQSISEYSGFSVARLLLNWNLDMHNWVIVRSTNKEHIRSN 529464

Query 297    AKASLYALSDPVRMILDRFQERVGTIRTINPTDFTRGGGSFFD 339
Sbjct 529465 AKASLYALSDPVRMILDRFQERVGTIRTINPTDFTRGGGSFFD 529593
```

[Link to Genome Browser](#), Strand = >

TcChr40-P

| organism=Trypanosoma cruzi_CL_Brener_Non-Esmeraldo-like
| version=2015-12-07 | length=2036760 | SO=chromosome
Length=2036760

Score = 149 bits (377), Expect = 2e-38, Method: Compositional matrix adjust.
Identities = 97/315 (31%), Positives = 150/315 (48%), Gaps = 45/315 (14%)
Frame = +2

```

Query 23      LPLRNGNSIPQCGFGTYRMTPTVAGA-AVEYAIHCGFRHIDCAKAYDNQNAIGEALQRFI 81
              + L N  +PQ G G +R      A AV +AI G+RHID A Y+N+ +G+   I
Sbjct 486278  VTLHNSARMPQLGLGVWRAQDGAETANAVRWAIIEAGYRHDITACIYNNKGVGGQ----I 486445

Query 82      STGNLKRREELFLTSKLPDQHPHVEKACRETLAELRVDYLDLYLIHWPVWVNHSPHF 141
              + REE+++T+K+W +DQ      A + L ++Y+DLYLIHWP
Sbjct 486446  RESGVPREEVWTTKVNWDQGYEKTAAAFERSRELLGLEIDLYLIHWP----- 486595

Query 142     TDDEKYPKDANGLPAVDDSVKLIIDTWRAMCELVDRNLVRSVGLSNCSEKHINEVMSD 201
              K +DTW+A+ +L + VR++G+SN H+ E+ +
Sbjct 486596  -----GKKKFVDTWKALEKLYEEKKVRIGVSNFEPHHLTELFKSKKI 486724

Query 202     YAPVNVQIELHPALVQRDLLNFHGANQIVTAAYSPLGMPFRFTHPDYKGLLSDASLQ 261
              P+VNQ+ELHP QR L F + I A+SPLG R      G L + L I+
Sbjct 486725  -RPMNVQVELHPLFQQRTLREFCKQHIAITAWSPLGSGDR-----TGFLKNHVLGEIA 486883

Query 262     EYSGFSVARLLLLNWNLDMHNVIIVRSTNKEHIRSNAKASLYALSOPVVMILDRFQE--RV 319
              + S A++++ W++ V I +STNK I+ N + L++ +D E R+
Sbjct 486884  KKHNSPAQWVIRWDIQHGIVTIPKSTNKRIQENFNWDFKLTTEEEMRQIDELNEDKRI 487063

Query 320     GTIRTINPTDFTRGG 334
              G +P +F GG
Sbjct 487064  GG----HPDNFFPGG 487096

```

[Link to Genome Browser](#), Strand = >

TcChr26-P

| organism=Trypanosoma cruzi_CL_Brener_Non-Esmeraldo-like
| version=2015-12-07 | length=801422 | SO=chromosome
Length=801422

Score = 130 bits (328), Expect = 3e-32, Method: Compositional matrix adjust.
Identities = 93/294 (32%), Positives = 135/294 (46%), Gaps = 42/294 (14%)
Frame = -3

```

Query 23      LPL---RNGNSIPQCGFGTYRMTPTVAGAAVEYAIHCGFRHIDCAKAYDNQNAIGEALQ 79
              LPL G IP G GTY + + AV A+ GFR +D A Y N+ +G A
Sbjct 588651  LPLLESSGGMRIPHMVGTYELYGDESRIAVLAALRLGFRLVDTAAGYHNEEHVGAA--- 588481

Query 80      VISTGNLKRREELFLTSKLPDQHPHVEKACRETLAELRVDYLDLYLIHWPVWVNHSP 138
              I ++REELF+ K+ P VEK RE++ +LR+ Y D LIHWP
Sbjct 588480  -IEESGVQREELFVVKIAPKAMASEELVEKGIRESVRKLR IAYADCVL IHPGCGGLK- 588307

Query 139     HFKTDEKYPKDANGLPAVDDSVKLIIDTWRAMCELVDRNLVRSVGLSNCSEKHINEVMS 198
              P++A G A W+ M L ++R +G+SN + +H ++ +
Sbjct 588306  -----PEEAEGHRAARHR-----CWKVMQALKKEGIIRHLGVSNFAPRHFPHLVGN 588169

Query 199     GSLYA-----PVVNQIELHPALVQRDLLNFHGANQIVTAAYSPL--GMPFRFTHP 246
              + A PV+NQIELHP VQ++ +F I+ YSPL G HP
Sbjct 588168  AEEEHLFDAANPCRPVINQIELHPLCVQEQEACDFCRERDIIILQQYSPLAQGHAKLVEHP 587989

Query 247     DYKGLLSDASLQSISEYSGFSVARLLLLNWNLDMHNVIIVRSTNKEHIRSNAKAS 300
              ++ D + G++V +LL W L V+VRS +KEH+ N A+
Sbjct 587988  VLCAIVKDL-----FPGYTVQDVLMLMGLSQGFCVVVRSRSKEHLAQNWAAA 587848

```

[Link to Genome Browser](#), Strand = >

TcChr40-S

| organism=Trypanosoma cruzi_CL_Brener_Esmeraldo-like
| version=2015-12-07 | length=2036759 | SO=chromosome
Length=2036759

Score = 117 bits (292), Expect = 1e-27, Method: Compositional matrix adjust.
Identities = 80/285 (28%), Positives = 128/285 (45%), Gaps = 45/285 (16%)
Frame = +3

```
Query 36      FGYRMTP---TVAGAAVEYAIHCGFRHIDCAKAYDNQNAIGEALQRVISTGNLKREELF 92
           FG +R P   +G + A   RH   Y+N+ +G+   I   + REE++
Sbjct 486597  FGGHRTEPRRRMRSGGQLRQATATLIRHT----FYNNERGVGQG----IRESGVPREEVW 486752

Query 93      LTSKLVPTDQHPHVEKACRETAEELRVDYLDLYLIHWPVWVWVHSPHFKTDDKYPKDAN 152
           +T+K+W +DQ   A   +   L ++Y+DLYLIHWP
Sbjct 486753  VTTKVWNSDQGYEKTAAAFERSRELLGLEIDLYLIHWP----- 486869

Query 153     GLPAVDDSVKLIIDWRAMCELVDRNLVRSVGLSNCSEKHINEVMSDGLYAPVWVNIQLH 212
           K +DTW+A+ +L +   VR++G+SN   H+ E+   + P+VNQ+ELH
Sbjct 486870  -----GKKKFVDTWKALEKLYEEKKVRIGVSNFEPHHLTELFQSCKI-RPMVWVQVELH 487028

Query 213     PALVQRDLLNFHGANQIVTAAYSPLGMPFRFTHPDYKGLLSDASLQSISEYSGFSVARLL 272
           P   QR L F + I   A+SPLG R   G L + L I++   S A+++
Sbjct 487029  PLFQQRTRLEFCKQHIAITAWSPLGSGDR-----TGFLKNHVLGEIAKKNKSPAQVV 487190

Query 273     LNWNLDMHNWVIVRSTNKEHIRSNAKASLYALSDPVRMILDRFQE 317
           + W++   V I +S NK I+ N   + L++   +D E
Sbjct 487191  IRWDIQHGIVTIPKSANKGRIQENFMWDFKLTEEDMRQIDELNE 487325
```

[Link to Genome Browser](#), Strand = >

TcChr26-S

| organism=Trypanosoma cruzi_CL_Brener_Esmeraldo-like
| version=2015-12-07 | length=801422 | SO=chromosome
Length=801422

Score = 64.7 bits (156), Expect = 6e-11, Method: Compositional matrix adjust.
Identities = 36/99 (36%), Positives = 52/99 (53%), Gaps = 9/99 (9%)
Frame = -3

```
Query 204     PVVNIQLHPALVQRDLLNFHGANQIVTAAYSPL--GMPSRFTHPDYKGLLSDASLQSI 261
           PV+NQIELHP VQ+++ +F   I+ YSPL G   HP   ++ D
Sbjct 588660  PVINQIELHPLCVQQEVCFRCRERDIIILQQYSPLAQGHAKLVEHPVLCAIVKDL----- 588499

Query 262     EYSGFSVARLLLWNWLDHMWVIVRSTNKEHIRSNAKAS 300
           + G++V +LL W L   V+VRS +KEH+ N A+
Sbjct 588498  -FPGYTVQDVLMLWGLSQGFCVVVRSRSKEHLAQWAAA 588385
```

Lambda	K	H	a	alpha	
0.320	0.136	0.420	0.792	4.96	

Gapped					
Lambda	K	H	a	alpha	sigma
0.267	0.0410	0.140	1.90	42.6	43.6

Effective search space used: 5050823715

Database:

TcruziCLBrenerEsmeraldo-likeGenome
 Posted date: Sep 25, 2020 10:18 AM
 Number of letters in database: 32,529,070
 Number of sequences in database: 41

Database:

TcruziCLBrenerNon-Esmeraldo-likeGenome
 Posted date: Sep 24, 2020 12:57 PM
 Number of letters in database: 32,529,072
 Number of sequences in database: 41

Matrix: BLOSUM62

Gap Penalties: Existence: 11, Extension: 1
 Neighboring words threshold: 13
 Window for multiple hits: 40

A busca a nível genômico por cópias da AKR120 retornou seis sequências.

Fonte:

produzido pelas autoras utilizando ferramentas disponíveis no programa TriTrypDB - disponível em <https://tritrypdb.org/tritrypdb/>.

Figura S6 – Regiões de reconhecimento dos sgRNAs na sequência codificante do gene *TcNTR-1*.
 CLUSTAL W (1.83) multiple sequence alignment

```

TcCLB.506791.70 ATGAAGAGAAATGGCATAAAAAGAGGCCCTGTGGGATTCTTTAATTTTGTATTGGCGTTGG
TcCLB.510611.60 ATGAGGAGAAATGACATAAAAAGACGCCCTGTGGGATTCTTTAATTTCTATTGGCGTTGG

TcCLB.506791.70 AATCGAGAAAACCTTTTGAGGAATGTTTCTACATTTGCCGAAAATGGCAGGCACGTGATT
TcCLB.510611.60 AATCGAGAAAACCTTTTGAGGAATTTTCTGCATTTGTGGAAAATGGCAGGCACGTGATT

TcCLB.506791.70 GGTATGGATGGTCTGTGGAGGCAGGTAGTGAAAAGGACGTGGGGCGTGGGAATTCATAC
TcCLB.510611.60 GGTATGGATGGTCTGTGGAGGCAGGTAGTGAAAAGGACATGGGGCGTGGGAATTCATCC

TcCLB.506791.70 ATGCCTCCTATTTTTTCTCGATGCCGCCATCACCATCATCATCATCATTGCCGCTTGAT
TcCLB.510611.60 ATGCCTCCTTTTTTCTCGATGCCGCCAT---CATCATCATCATCATTGCCGCTGGAT

TcCLB.506791.70 ACCATGAAACGTGTAGTACACGAGCGTCGCTCCTGCAAGCGATTTGACCCAACAAAGCCA
TcCLB.510611.60 GCCATGAAACGTGTAGTACACGAGCGTCGCTCCTGCAAGCGATTTGACCCAACAAAGTCA

TcCLB.506791.70 ATTGATCTCGACGTAGTTTCTGATCTTCTCGCCATGACGGTGCGGGCTCCTACCGCCCTG
TcCLB.510611.60 ATTGACCTCGACGTGGTTTCTGATCTTCTCGCCATGACGGTACGGGCTCCTACCGCCCTG

TcCLB.506791.70 AACTTACAGCCGTGGGTGGCTGTTGTGCATACACGAGGAGGAGCAAAGAGAAAACCTCTCG
TcCLB.510611.60 AACTTACAGCCGTGGGTGGCTGTTGTGCATACACGAGGAGGAGCAACGAGAAAACCTCTCG

TcCLB.506791.70 CGTGCGGCCCTGGGGCAACCCGACGCTCGTGACGCACCGGTGACTGTTGTGTTTCGCTGGT
TcCLB.510611.60 CGTGCGGCCCTGGGGCAACCCGACGCTCGTGACGCACCGGTGACAGTTGTGTTTCGCTGGT

TcCLB.506791.70 GACATGGAGCCGGAGAGCAATGCCCCGGCCGCGCTTGAGATGGGGCTGGAGAGTGGGTAC
TcCLB.510611.60 GACATGGAGCCGGAGAGCAATGCCCCGGCCGCGCTTGAGATGGGGCTGGAGAGTGGGTAC

TcCLB.506791.70 TACCACTCACTTTATGGCGCAGCCTATCTGCGACATGCGTACTACTACTTCACGGTGGC
TcCLB.510611.60 TACCACTCACTTTATGGCGCAGCCTATCTGCGACATGCGTACTACTACTTCACGGTGGC

TcCLB.506791.70 CCATGTGAAGCCATGTCACATGTTAAGGCGATTGTGTCTGCATGGTATAGTGAAAGCACT
TcCLB.510611.60 CCATGTGAAGTCATGTCACATGTCAAGGCGATTGTGTCTGCATGGTATAGTGAAAGCACT

TcCLB.506791.70 GGCAACGCGATGCTGTGGTTCTAGAAATAAGCAGGCGTATGCGTGGAAGCAGGTTATG
TcCLB.510611.60 GGCAACGCGATGCTGTGGTTCTAGAAATAAGCAGGCGTATGCGTGGAAGCAGGTTATG
  
```

Alinhamento global das sequências de NTR-1 de CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like por *Pairwise Sequence Alignment* usando o Needle (EMBOSS). Também estão indicadas as regiões de reconhecimento dos sgRNA escolhidos: azul é a região de pareamento do sgRNA NTR1_585_revcom; vermelho é a região de pareamento do sgRNA NTR1_375_revcom_FW ; roxo é a região de pareamento do sgRNA NTR1_565_revcom_FW.

Fonte:
 produzido pelas autoras.

Figura S7 – Regiões de reconhecimento dos sgRNAs na sequência codificante do gene *TcADH*.

CLUSTAL W (1.83) multiple sequence alignment

```

TcCLB.511277.60 ATGTTTCGCTTCTCACGCCCAAGTGTGCGGCCCTCCACCTTTTTCATGCCGAAGGTCAGC
TcCLB.506357.50 ATGTTTCGCTTCTCACGCCCAAGTGTGCGGCCCTCCACCTTTTTCATGCCAAAGGTCAGT

TcCLB.511277.60 TACCTTGGCGTTGGGGCGCTCGACAGTGCACGAGCAAGATGACAAAGTTGGGCTTTAAG
TcCLB.506357.50 TACCTTGGCGTTGGGGCGCTCGACAGTGCACGAGCAAGATGACAAAGTTGGGCTTTAAG

TcCLB.511277.60 AAGCCCTTATTGGTGTGCGACACCGGCATCGAAAAGGCAAGCATTCTTGAACCGTCATG
TcCLB.506357.50 AAGCCCTTATTGGTGTGCGACACCGGCATCGAAAAGGCAAGCATTCTTGAACCGTCATG

TcCLB.511277.60 GATGTGCTTAAGAAAAAGTGTCAATGGACCTGTCTGATTCAAGGATGTGCATCCGAAC
TcCLB.506357.50 GATGTGCTTAAGAAAAAGTGTCAATGGACCTGTCTGATTCAAGGATGTGCATCCGAAC

TcCLB.511277.60 CCAACTGCCACAACTGGAGACTGGTCTGGAGTTGCTGAACAAGAATGCGTGCATTCT
TcCLB.506357.50 CCAACTGCCACAACTGGAGACTGGTCTGGAGTTGCTGAACAAGAATGCGTGCATTCT

TcCLB.511277.60 ATCATCGCTATTGGAGTGGATCGCCGACAGGACTGTGCAAGGGCATTGCACCTGTCAAG
TcCLB.506357.50 ATCATCGCTATTGGAGTGGATCGCCGACAGGACTGTGCAAGGGCATTGCACCTGTCAAG

TcCLB.511277.60 GC TAATGGGGG ACCCATCTATGACTACGAGGGTCTGGACAAGGCTGAGAAGGACAGTAC
TcCLB.506357.50 GC TAATGGGGG ACCCATCTATGACTACGAGGGTCTGGACAAGGCTGAGAAGGACAGTAC

TcCLB.511277.60 CCGCTGGTTGCATAAACACAACATCGGGCACAGCAAGCGAAATAACGCGCTTTGCGGTT
TcCLB.506357.50 CCGCTGGTTGCATAAACACAACATCGGGCACAGCAAGCGAAATAACGCGCTTTGCGGTT

TcCLB.511277.60 ATTACGGACGAGAAGCGACACGTCGAAGTGGTATTACGACAGATACTGTGACTCCACTC
TcCLB.506357.50 ATTACGGACGAGAAGCGACACGTCGAAGTGGTATTACGACAGATACTGTGACTCCACTC

TcCLB.511277.60 GTTGCCATAGACGACGCCAACTGATGATGGGACAGCCGCTTCTCTTACTGCTGCCACA
TcCLB.506357.50 GTTGCCATAGACGACGCCAACTGATGATGGGACAGCCGCTTCTCTTACTGCTGCCACG

TcCLB.511277.60 GGTATGGATGCTCTCACACATGCAAGTGAAGCGTATGTTTCAAGTGGCGCAAGCGAAACC
TcCLB.506357.50 GGTATGGATGCTCTCACACATGCAAGTGAAGCGTATGTTTCAAGTGGCGCAAGCGAAACC

TcCLB.511277.60 ACAGATGCGTGCAGCTATATGCCATGAGACTTATCAACAGCATCTTTCTAATGCGGTG
TcCLB.506357.50 ACAGATGCGTGCAGCTATATGCCATGAGACTTATCAACAGCATCTTTCTAATGCGGTG

TcCLB.511277.60 AAGGATGGGAAGAATCTTGTGGCGGAGAGGGTATGTGCTACGCGCAGCATCTTGTCTGGA
TcCLB.506357.50 AAGGATGGCAAGAATCTTGTGGCACGAGAGGGTATGTGCTACGCGCAGCATCTTGTCTGGA

TcCLB.511277.60 ATGGCATTAAACAGTGCAGGCTCGGTTACGTGCACGCAATGGCTCATCAGCTTGGTGGC
TcCLB.506357.50 ATGGCATTAAACAGTGCAGGCTCGGTTACGTGCACGCAATGGCTCATCAGCTTGGTGGC

TcCLB.511277.60 TTCTACGACCTCCCCATGGCGTCTGTAAACGCTATCCTGCTGCCACAGTGGAAAGTTAC
TcCLB.506357.50 TTCTACGACCTCCCCATGGCGTCTGTAAACGCTATCCTGCTGCCACAGTGGAAAGTTAC

TcCLB.511277.60 AACTCTTTGATATGTGCGCGCCGTTTGGGTGATGTTGCCGTGCAATTGGGTGTCAGTACG
TcCLB.506357.50 AACTCTTTGATATGTGCGCGCCGTTTGGGTGATGTTGCCGTGCAATTGGGTGTCAGTACG

TcCLB.511277.60 ACAGGCATGGATGATAANGAATGCCGCTGCTGCAATTGATGCGATTGCTGCTTTATCG
TcCLB.506357.50 ACAGGCATGGACGATAANGAATGCCGCTGCTGCAATTGATGCGATTGCTGCTTTATCA

TcCLB.511277.60 AAGGAAGTTCAAATCCCACAGGATTTGAACAAC TTGGTATGAAGGAGAAGACATTCCC
TcCLB.506357.50 AAGGAAGTTCAAATCCCACAGGATTTGAACAAC TTGGTATGAAGGAGAAGGACATTCCC

TcCLB.511277.60 GCGCTGGCAGAATCAGCCTTGAAGGATGTTTGCGCCGCAACCCGCGACAGGGTAGC
TcCLB.506357.50 GCGCTGGCAGAATCAGCCTTGAAGGATGTTTGCGCCGCAACCCGCGACAGGGTAGC

TcCLB.511277.60 AAGGAAGATGTGATGAAGATCTACCGTGAGTCAATGTAG
TcCLB.506357.50 AAGGAAGATGTGATGAAGATCTACCGTGAGTCAATGTAG

```

Alinhamento global das sequências de *TcADH* de CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo-like por *Pairwise Sequence Alignment* usando o Needle (EMBOSS). Também estão indicadas as regiões de reconhecimento dos sgRNA escolhidos: azul é a região de pareamento do sgRNA ADH_349; vermelho é a região de pareamento do sgRNA ADH_456; roxo é a região de pareamento do ADH_204.

Fonte:

produzido pelas autoras.

Figura S8 – Regiões de reconhecimento dos sgRNAs na sequência codificante do gene *TcAKR49*.

CLUSTAL W (1.83) multiple sequence alignment

```

TcCLB.511287.49 --ATGAATTGCAATTACAACGTGTGACACTCCATAACTCCGCGCAATGCCGCGAGCTTG
TcCLB.506213.50 TCATGAATTGCAATTACAGCTGTGTGACACTCCATAACTCCGTGCGAATGCCGCGAGCTTG

TcCLB.511287.49 GCTTGGGTGTTTGGCGGSCACAGGACGGAGCCGAGACGGCGAATGCGGTCCGGTGGGCAA
TcCLB.506213.50 GCTTGGGTGTTTGGCGGSCACAGGACGGAGCCGAGACGGCGAATGCGGTCCGGTGGGCAA

TcCLB.511287.49 TTGAGGCAGGCTACCGCCACATCGACACGGCATGC-ATCTACAACAACGAAAAGGGCGTG
TcCLB.506213.50 TTGAGGCAGGCTACCGCCACATTGATACGGCATACTTCTACAACAACGAAAAGGGCGTG

TcCLB.511287.49 GGGCAGGGGATCCGCGAATCGGGAGTGCCCGGTGAGGAGGTGGGTAACAACAAAAGTA
TcCLB.506213.50 GGGCAGGGGATCCGCGAATCGGGAGTGCCCGGTGAGGAGGTGGGTAACAACAAAAGTA

TcCLB.511287.49 TGGAAATCAGATCAAGGATACGAAAAGACGCTTGCCTTTGAACGCAGCCGTGAGCTG
TcCLB.506213.50 TGGAAATCAGATCAAGGATACGAAAAGACGCTTGCCTTTGAACGCAGCCGTGAGCTG

TcCLB.511287.49 CTCGGATTGGAGTACATTGATCTCTACTTAATTCACCTGGCCGGGAAGAAGAAGTTTGT
TcCLB.506213.50 CTCGGACTGGAGTACATTGATCTCTACTTAATTCACCTGGCCGGGAAGAAGAAGTTTGT

TcCLB.511287.49 GACACATGGAAAGCATTGGAGAAGCTCTACGAGGAGAAGAAGGTGCGGCCATTGGCGTT
TcCLB.506213.50 GACACATGGAAAGCATTGGAGAAGCTCTACGAGGAGAAGAAGGTGCGGCCATTGGCGTT

TcCLB.511287.49 TCCAACCTTTGAGCCACACCACCTCACGGAGCTTTTTAAAAGCTGCAAAATTCGGCCGATG
TcCLB.506213.50 TCCAACCTTTGAGCCACACCACCTCACGGAGCTTTTTCAAAGCTGCAAAATTCGGCCGATG

TcCLB.511287.49 GTCAACCAAGTGGAGCTGCACCCGCTGTTCCAACAGCGTACTTTGCGGGAGTTCTGCAAG
TcCLB.506213.50 GTGAACCAAGTGGAGCTGCACCCGCTGTTCCAACAGCGTACTTTGCGGGAGTTCTGCAAG

TcCLB.511287.49 CAACACAACATTGCCATCACAGCCTGGTCTCCGCTTGGCAGCGGGGACCGGACGGGTTTT
TcCLB.506213.50 CAACACAACATTGCCATCACAGCCTGGTCTCCGCTTGGCAGCGGGGACCGGACGGGTTTT

TcCLB.511287.49 CTGAAGAATCACGTGCTGGGGGAGATTGCCAAGAAGCACAAATAATCCCGCGCAGGTT
TcCLB.506213.50 CTGAAGAATCACGTGCTGGGGGAGATTGCCAAGAAGCACAAATAATCCCGCTCAGGTT

TcCLB.511287.49 GTCATCCGCTGGGACATTGACACGGTATCGTGACAATTCCTCAAGTCCACAATAAGGGG
TcCLB.506213.50 GTCATCCGCTGGGACATTCAACACGGTATCGTGACAATTCCTCAAGTCCACAATAAGGGG

TcCLB.511287.49 CGCATTCAAGAAAACCTTAATGTTTGGGACTTTAACTTACGGAAGAGGAGATGCGTCAA
TcCLB.506213.50 CGCATTCAAGAAAACCTTAATGTTTGGGACTTTAAGCTTACGGAAGAGGACATGCGTCAA

TcCLB.511287.49 ATCGATGAACTCAACGAGGATAAACGCATCGGCGGACACCCGATAATTTTTCCCTGGT
TcCLB.506213.50 ATCGATGAACTCAACGAGGATAAAGCGCTTCGGCGGACACCCGATAATTTTTCCCTGAT

TcCLB.511287.49 GGAGAGGAGTGA
TcCLB.506213.50 GGAGAGGAGTGA

```

Alinhamento global das sequências de *TcAKR49* de CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo like por Pairwise Sequence Alignment usando o Needle (EMBOSS). Também estão indicadas as regiões de reconhecimento dos sgRNA escolhidos: azul é a região de pareamento do sgRNA AKR_53; vermelho é a região de pareamento do sgRNA AKR49_252 ; roxo é a região de pareamento do sgRNA AKR49_229_revcom. Fonte: produzido pelas autoras.

Figura S9 – Regiões de reconhecimento dos sgRNAs na sequência codificante do gene *TcAKR120*.

CLUSTAL W (1.83) multiple sequence alignment

```

TcCLB.511627.12 -ATGCCCGGTGCGAACGTTCTGAGTTTCAAGCCACTCGGGGGGCCGTGCTCGGCGTGTC
TcCLB.507801.11 AATGCCCGGTGCGAACGTTCTGAGTTTCAAGCCACTCGGGGGGCCGTGCTCGGCGTGTC

TcCLB.511627.12 GCATCGGCTACCGTTGAGAAATGGAATAGCATTCTCAGTGC GGCTTTGGAACGTACCG
TcCLB.507801.11 GCATCGGCTGCGTTGAGAAATGGAATAGCATTCTCAGTGC GGCTTTGGAACGTACCG

TcCLB.511627.12 CATGACCCCCACAGTGGCAGGAGCTGCCGTGGAATATGCCATACATTGTGGTTTTCGACA
TcCLB.507801.11 CATGAATCCACAGTGGCAGGAGCTGCCGTGGAATATGCCATACATTGTGGTTTTCGACA

TcCLB.511627.12 TATTGATTGTGCTAAGGCATACGACAACCAAGACGCTATTGGAGAGGCACCTCAACGTGT
TcCLB.507801.11 TATTGATTGTGCTAAGGCATACGACAACCAAGACGCTATTGGAGAGGCACCTCAACGTGT

TcCLB.511627.12 GATATCAACTGGCAATTTGAAGCGTGAGGAATGTTTTTGACGTCAAAGCTTTGGCCAC
TcCLB.507801.11 GATATCAACTGGCAAGTTGAAGCGTGAGGAATGTTTTTGACGTCAAAGCTTTGGCCAC

TcCLB.511627.12 AGATCAACACCCAATTACAGTGGAGAAGGCTTGTGAGAGACGTTAGCGGAGCTGCGTGT
TcCLB.507801.11 AGATCAACACCCAATTACAGTGGAGAAGGCTTGTGAGAGACGTTAGCGGAGCTGCGTGT

TcCLB.511627.12 GGACTACCTTGACTTGATCTTATTCACTGGCCAGTGGTGTGGAATCATTCCCCGCAATT
TcCLB.507801.11 GGACTACCTTGACTTGATCTTATTCACTGGCCAGTGGTGTGGAATCATTCCCCGCAATT

TcCLB.511627.12 CAAGACCGATGACGAGAAATACCCCAAGACGCCAACGGACTTC TGC GGTTGATGATAG
TcCLB.507801.11 AAAGACCGATGACGAGAAATACCCCAAGACGCCAACGGACTTC TGC GGTTGATGATAG

TcCLB.511627.12 TGTCAAATTCGATACTTGGAGGGCTATGTGTGAAC TTGTGATAGAAATTTGGTGCG
TcCLB.507801.11 TGTCAAATTCGATACTTGGAGGGCTATGTGTGAAC TTGTGATAGAAATTTGGTGCG

TcCLB.511627.12 TTCAGTGGGCTCTCGAACTGTAGTGAGAAACATATAACGAAGTATGAGTGATGGAAG
TcCLB.507801.11 -TCAGTGGGCTCTCGAACTGTAGTGAGAAACATATAACGAAGTATGAGTGATGGAAG

TcCLB.511627.12 TCTGTATGCACCGGTGGTAAACAGATTGAACTTCAACCGGCGTTGGTCCAGCGTGACCT
TcCLB.507801.11 TCTGTATGCACCGGTGGTAAACAGATTGAACTTCAACCGGCGTTGGTCCAGCGTGACCT

TcCLB.511627.12 TCTTAATTTCCACGGTGCAAAATCAATAGTGCAGCTGCTTACAGTCTCTGGGGATGCC
TcCLB.507801.11 TCTTAATTTCCACGGTGCAAAATCAATAGTGCAGCTGCTTACAGTCTCTGGGGATGCC

TcCLB.511627.12 TTCAAGATTACACACCCGGACTACAAGGGTTGTTGTCCGACGCATATTACAATCCAT
TcCLB.507801.11 TTCAAGATTACACACCCGGACTACAAGGGTTGTTGTCCGACGCATATTACAATCCAT

TcCLB.511627.12 TTCCGAGATTACAGGATTAGCGTGGCTCGTTACTACTAACTGGAACCTTGACATGCA
TcCLB.507801.11 TTCCGAGATTACAGGATTAGCGTGGCTCGTTACTACTAACTGGAACCTTGACATGCA

TcCLB.511627.12 CAATGTCGTGATTGTGCGTTCAACCAACAAGGAACACATCAGATCGAACGCCAAGGCCCTC
TcCLB.507801.11 CAATGTCGTGATTGTGCGTTCAACCAACAAGGAACACATCAGATCGAACGCCAAGGCCCTC

TcCLB.511627.12 TCTTTATGCACTCTCGGATCCCCTGAGAAATGATCCTGGACCGATTCCAGGAGCGTGGG
TcCLB.507801.11 TCTTTATGCACTCTCGGATCCCCTGAGAAATGATCCTGGACCGATTCCAGGAGCGTGGG

TcCLB.511627.12 AACAAATCGTACAATCAATCCGACGGACTTCACTCGTGGCGGAGGGAGCTTCTTTGACTG
TcCLB.507801.11 AACAAATCGTACAATCAATCCGACGGACTTCACTCGTGGCGGAGGGAGCTTCTTTGACTG

TcCLB.511627.12 A
TcCLB.507801.11 A

```

Alinhamento global das sequências de *TcAKR120* de CL Brener Non-Esmeraldo-like e CL Brener Esmeraldo like por Pairwise Sequence Alignment usando o Needle (EMBOSS). Também estão indicadas as regiões de reconhecimento dos sgRNA escolhidos: azul é a região de pareamento do sgRNA AKR120_692; vermelho é a região de pareamento do sgRNA AKR120_447; roxo é a região de pareamento do sgRNA AKR120_463_revcom. Fonte: produzido pelas autoras.