



MINISTÉRIO DA EDUCAÇÃO

UTFPR – UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

CAMPUS CORNÉLIO PROCÓPIO

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA



Viviane Aparecida Gobetti

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE RNAs NÃO CODIFICADORES EM
ISOLADOS DE *Bacillus thuringiensis* (*Bacillus cereus sensu lato*) POR
SEQUENCIAMENTO DO CONJUNTO DE PEQUENOS RNAs**

DISSERTAÇÃO

CORNÉLIO PROCÓPIO

2021

VIVIANE APARECIDA GOBETTI

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE RNAs NÃO CODIFICADORES EM
ISOLADOS DE *Bacillus thuringiensis* (*Bacillus cereus sensu lato*) POR
SEQUENCIAMENTO DO CONJUNTO DE PEQUENOS**

*Identification and characterization of noncoding RNAs in Bacillus thuringiensis
(Bacillus cereus sensu lato) isolates by sequencing the set of smalls*

Dissertação apresentada como requisito para
obtenção do título de Mestre em Bioinformática da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Orientador: Dr. Laurival Antônio Vilas-Boas.

Coorientador: Dr. André Yoshiaki Kashiwabara.

CORNÉLIO PROCÓPIO

2021



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite o download e o compartilhamento da obra desde que sejam atribuídos créditos ao(s) autor(es), sem a possibilidade de alterá-la ou utilizá-la para fins comerciais.



VIVIANE APARECIDA GOBETTI

**IDENTIFICAÇÃO E CARACTERIZAÇÃO DE RNAs NÃO CODIFICADORES EM
ISOLADOS DE BACILLUS THURINGIENSIS (BACILLUS CEREUS SENSU LATO) POR
SEQUENCIAMENTO DO CONJUNTO DE PEQUENOS RNAs**

Trabalho de pesquisa de mestrado
apresentado como requisito para obtenção
do título de Mestra Em Bioinformática da
Universidade Tecnológica Federal do Paraná
(UTFPR). Área de concentração:
Bioinformática.

Data de aprovação: 14 de
Dezembro de 2020

Prof Laurival Antonio Vilas Boas, - Universidade Tecnológica Federal do Paraná

Prof Ivan Rodrigo Wolf, Doutorado - Universidade Estadual Paulista Júlio de

Mesquita Filho (Unesp) Prof Luiz Filipe Protasio Pereira, - Universidade

Tecnológica Federal do Paraná

Dedico este trabalho a Deus e
aos meus pais.

AGRADECIMENTOS

Agradeço primeiramente a Deus por se fazer sempre presente, por sempre colocar no meu caminho pessoas especiais. Por me abençoar, guiar e me dar forças e coragem para alcançar todos os meus objetivos.

Aos meus pais Celso e Leonilda, por sempre me apoiarem em minhas escolhas e me incentivarem a lutar por todos os meus sonhos, por estarem sempre ao meu lado e por torcerem sempre por mim. Amo muito vocês!

Ao professor Laurival Antônio Vilas-Boas por aceitar ser meu orientador e por acreditar em mim para desenvolver este trabalho, agradeço por todo conhecimento transmitido, pela paciência e disponibilidade que sempre teve do início até a conclusão deste trabalho. Muito obrigada!

Ao professor André Yoshiaki Kashiwabara pela co-orientação, pelas dicas e sugestões que contribuíram de forma significativa para este trabalho.

Agradeço de forma especial a Kátia Brumatti Gonçalves por todo suporte prestado, pelo conhecimento compartilhado e pela paciência em me ajudar no início do projeto, pois foi de imensurável importância para o desenvolvimento deste trabalho.

Agradeço a todas as pessoas que me ajudaram até aqui, direta ou indiretamente para que este trabalho fosse realizado. Por todas as pessoas com quem tive o prazer de conviver durante este período do mestrado, das amigadas que conquistei em cada disciplina cursada e de cada pessoa que conheci no decorrer do caminho, pois foram de grande importância para tornar a caminhada até aqui ainda mais especial.

*“Um navio está seguro quando está no porto,
mas não é para isso que se fazem navios”*

William Shedd

Lista de Figuras

Figura 1: Árvore filogenética do grupo <i>Bacillus cereus</i> (Extraído de RASKO et al, 2006).....	6
Figura 2: Duas classes de RNAs - os codificadores e não codificadores (Adaptado de KALANTIDIS, 2015).....	11
Figura 3: Estrutura secundária de um tRNA. Extraído de [VIEIRA, 2016]	13
Figura 4: Estrutura secundária do rRNA. Extraído de Rfam.....	13
Figura 5: Gráfico representando a curva de crescimento de uma bactéria. (T-0) Final da fase logarítmica de crescimento, (T-1) Início de fase estacionária. (Autoria própria).	21
Figura 6: Visualizador Software Artemis - exemplo de região selecionada para análise de ncRNAs	22
Figura 7: Fluxograma das etapas do projeto.....	23
Figura 8: Gráfico da relação de candidatos para as famílias mais freqüentes nas amostras.....	26
Figura 9: Estrutura secundária proposta para 5_8S_rRNA	27
Figura 10: Estrutura secundária SSU_rRNA_bacteria	28
Figura 11: Estrutura secundária de PyrR.....	29
Figura 12: Estrutura secundária de L10_leader	30
Figura 13: Comparação dos resultados das famílias de ncRNAs	36
Figura 14: Abrir arquivo GenBank (.gbff) no Artemis.....	48
Figura 15: Selecionar arquivo GenBank (.gbff) no Artemis	49
Figura 16: Abrir arquivo de alinhamento .sorted.bam.....	49
Figura 17: Janela de visualização do genoma no Artemis	50
Figura 18: Extrair sequência FASTA no Artemis.....	50
Figura 19: Exemplo de sequência FASTA extraída com Artemis.....	51
Figura 20: Arquivo multiFASTA com todas as sequências da Amostra 1	51

Lista de Tabelas

Tabela 1: Ferramentas utilizadas no projeto	17
Tabela 2: Linhagem Bt 407 e as identificações e os respectivos códigos de acesso das sequências do genoma	19
Tabela 3: Relação das famílias de ncRNAs encontrados. Novas famílias atribuídas aos resultados após análise posterior dos limites de inclusão apresentam-se marcadas com (*).	24
Tabela 4: Famílias de ncRNAs exclusiva de cada amostra analisada	26
Tabela 5: Comparação dos resultados com outras metodologias.....	32
Tabela 8: Possíveis sequências candidatas para Amostra 1	52
Tabela 9: Possíveis sequências candidatas para Amostra 2	55
Tabela 10: Relação entre as amostras	59
Tabela 11: Sequências não significativas resgatadas em novo limite de inclusão - Amostra 1	59
Tabela 12: Sequências não significativas resgatadas em novo limite de inclusão - Amostra 2.....	59

Lista de Abreviaturas e Siglas

Bt	<i>Bacillus thuringiensis</i>
CM	Modelo de Covariância
Cry	Proteínas cristalinas
CTN	Cationic Amino Acid transporter
Cyt	Proteínas citolíticas
DNA	Ácido desoxirribonucléico
EMBL	<i>European Molecular Biology Laboratory</i>
FTP	<i>File Transfer Protocol</i>
Infernal	Inferência de alinhamento de RNA (<i>INFERENCE of RNA Alingment</i>)
Kb	Quilobase
Mb	Megabase
miRNA	micro RNA
mL	Mililitro
NCBI	Centro Nacional de Informações em Biotecnologia (<i>National Center for Biotechnology Information</i>)
ncRNA	Ácido ribonucléico não codificador
NRDR	<i>Non-coding RNA Databases Resource</i>
ORF	Quadro aberto de leitura
pb	pares de bases
PCR	Reação em cadeia de polimerase
piRNA	Piwi-interacting RNA
RNA	Ácido Ribonucléico
rpm	rotação por minuto
snoRNAs	Small nucleolar RNAs
siRNA	Small Interfering RNA
sRNA	Pequeno RNA não mensageiro (<i>small non-messenger RNA</i>)
UEL	Universidade Estadual de Londrina
UTFPR	Universidade Tecnológica Federal do Paraná
Xist	X-interacting Specific transcript

Resumo

O grupo *Bacillus cereus* agrega oito espécies bacterianas que estão intimamente relacionadas entre si, no entanto, suas relações filogenéticas e taxonômicas ainda são debatidas. As três principais espécies do grupo são *Bacillus anthracis*, *Bacillus cereus* e *Bacillus thuringiensis*. Evidências indicam uma estreita semelhança nos genomas destas três bactérias, propondo que sejam consideradas como sendo da mesma espécie. A distinção entre as espécies do grupo *B. cereus* é atribuída essencialmente às características fenotípicas e às propriedades patogênicas, cujos genes que as codificam, possuem muito de suas especificidades atribuída aos seus plasmídeos. Apesar da semelhança, fenotipicamente são distintas e apresentam particularidades importantes. *B. anthracis* é um patógeno para humanos e outros animais, *B. cereus* é um importante contaminante de alimentos e *B. thuringiensis* é a bactéria mais utilizada no controle biológico de pragas e vetores de doenças. Desta forma, este trabalho visa contribuir ao esclarecimento sobre a proximidade genética entre estes três constituintes do grupo do *Bacillus cereus* sensu lato mediante identificação e caracterização de RNAs não codificadores na espécie de *B. thuringiensis*, através da identificação e caracterização de RNAs não codificadores na linhagem Bt407 de *Bacillus thuringiensis*, agregando a identificação e caracterização de sequências homólogas no banco de dados Rfam. A metodologia proposta neste trabalho permitiu a identificação de um total de 355 sequências candidatas a ncRNAs distribuídas em 49 famílias. Com a remoção dos RNAs comuns (tRNA, tmRNA, 58s_RNA e 5s_RNA), a metodologia identificou 223 sequências distribuídas em 45 famílias. O presente trabalho ainda apresenta uma comparação da metodologia de busca por ncRNAs proposta com outros métodos de busca realizados por outros autores para linhagem Bt407 de *Bacillus thuringiensis*.

Palavras-chave: *Bacillus thuringiensis*, *Bacillus cereus*, RNAs não codificadores.

Abstract

The Bacillus cereus group aggregates eight bacterial species that are closely related to each other, however, their phylogenetic and taxonomic relationships are still debated. The Bacillus anthracis, Bacillus cereus and Bacillus thuringiensis are the three main species in the cereus group. There is a close genomic similarity between these three bacteria, which leads to the proposal that they can be considered as the same species. The distinction between species of the B. cereus group is essentially attributed to phenotypic characteristics and pathogenic properties, whose genes encoding them, have much of their specificities attributed to their plasmids. Despite the similarity, they are phenotypically distinct and have important particularities. B. anthracis is a pathogen for humans and other animals, B. cereus is an important food contaminant and B. thuringiensis is the most used bacterium in the biological control of pests and disease vectors. Thus, this work aims to contribute to the clarification on the genetic proximity between these three constituents of the Bacillus cereus sensu lato group through the identification and characterization of non-coding RNAs in the species of B. thuringiensis, through the identification and characterization of non-coding RNAs in the lineage Bacillus thuringiensis bt407, adding the identification and characterization of homologous sequences in the Rfam database. The methodology proposed in this work allowed the identification of a total of 355 candidate sequences for ncRNAs distributed in 49 families. With the removal of common RNAs (tRNA, tmRNA, 58s_RNA and 5s_RNA), the methodology identified 223 sequences distributed in 45 families. The present work also presents a comparison of the proposed ncRNAs search methodology with other search methods carried out by other authors for Bacillus thuringiensis strain Bt407.

Keywords: *Bacillus thuringiensis, Bacillus cereus, non-coding RNAs.*

Sumário

LISTA DE FIGURAS	I
LISTA DE TABELAS	II
LISTA DE ABREVIATURAS E SIGLAS	III
RESUMO	IV
ABSTRACT	V
1. INTRODUÇÃO	2
2. JUSTIFICATIVA	3
2.1 REVISÃO BIBLIOGRÁFICA	4
1.1 CARACTERÍSTICAS DO GRUPO <i>BACILLUS CEREUS</i>	4
1.1.1 <i>Bacillus anthracis</i>	6
1.1.2 <i>Bacillus cereus</i>	7
1.1.3 <i>Bacillus thuringiensis</i>	7
1.1.4 Genômica de <i>Bacillus thuringiensis</i>	8
1.2 RNAs NÃO CODIFICANTES.....	9
1.3 BIOINFORMÁTICA NA BUSCA POR NCRNAS.....	14
3. OBJETIVO GERAL	16
3.1 OBJETIVOS ESPECÍFICOS.....	16
4. MATERIAIS E MÉTODOS	17
4.1 FERRAMENTAS COMPUTACIONAIS.....	17
4.1.1 <i>Infernal/Rfam</i>	17
4.1.2 <i>Artemis</i>	18
4.1.3 <i>Bowtie</i>	18
4.1.4 <i>RNAfold</i>	19
4.2 ESTRATÉGIAS PARA BUSCA E CARACTERIZAÇÃO DE NCRNAS.....	19
4.2.1 <i>Mineração de dados</i>	19
4.2.2 <i>Extração de sequências de RNA</i>	20
4.2.3 <i>Anotação</i>	22
4.4 FLUXOGRAMA DO PROJETO.....	23
5. RESULTADOS E DISCUSSÕES	23
5.1 PREDIÇÃO E AVALIAÇÃO DOS NCRNAS IDENTIFICADOS.....	23
5.2 FAMÍLIAS DE NCRNAS COM MAIOR NÚMERO DE CANDIDATOS.....	27
5.3 COMPARAÇÃO DE METODOLOGIAS.....	30
5.3.1 <i>Metodologia proposta por Appel, 2018. Predição, análise e validação de RNAs não codificadores em Bacillus thuringiensis</i>	31
5.3.2 <i>Metodologia proposta por Gonçalves, 2019</i>	31
5.3.3 <i>Comparando os resultados</i>	32
6. CONCLUSÕES	36
7. REFERÊNCIAS BIBLIOGRÁFICAS	37

1. INTRODUÇÃO

O grupo *Bacillus cereus*, também conhecido como *B. cereus sensu lato*, consiste de bactérias aeróbicas Gram-positivas, em forma de bastonete, formadoras de esporos, que são difundidas em ambientes naturais [LIU et. al., 2015]. O grupo do *B. cereus* agrega oito espécies bacterianas que estão intimamente relacionadas entre si, porém, suas relações filogenéticas e taxonômicas ainda são debatidas. As três principais espécies do grupo são *Bacillus anthracis*, *Bacillus cereus* e *Bacillus thuringiensis*, que foram reconhecidas e estabelecidas no início dos anos de 1900, sendo cada uma delas exibiu traços fenotípicos distintos [OKINAKA, 2012]. Evidências indicam uma estreita semelhança na sequência de DNA presente nos genomas destas três bactérias mencionadas, sendo que alguns autores sugerem que elas devem ser consideradas como sendo da mesma espécie [HELGASON et. al., 2000; ROSSI et. al, 2018].

Em relação as três espécies principais do grupo *cereus*, *B. anthracis* e *B. cereus* são as únicas espécies do grupo patogênicas para humanos, enquanto que *B. thuringiensis* é patogênica para insetos, sendo utilizada como agente de controle biológico [OKINAKA, 2012]. As pequenas diferenças cromossômicas encontradas entre as três principais espécies do grupo são principalmente por causa de genes que codificam características metabólicas, que permitem o uso de diferentes fontes de nutrientes, mas, que não são suficientes para explicar a aquisição de diferentes nichos ecológicos [VILAS-BÔAS et al., 2007].

A distinção entre as espécies do grupo *B. cereus* é atribuída essencialmente às características fenotípicas e às propriedades patogênicas, cujos genes que as codificam, possuem muito de sua especificidade atribuída aos seus plasmídeos, que variam em tamanho e número [CHAVES, 2015]. Os cromossomos deste grupo, apresentam um alto nível de sintonia com diferenças limitadas no conteúdo genético, o que torna questionáveis as interpretações sobre a especiação de seus membros [NADAL, 2015].

No entanto, novos esforços ainda são necessários para demonstrar a origem das diferenças entre os membros do grupo de *B. cereus* [ROSSI et. al., 2018]. Algumas dessas diferenças podem ter resultado de divergência de sequências ou conteúdo genético em cromossomos e plasmídeos [VILAS-BÔAS et al., 2007].

Entretanto, diferenças na patogenicidade e, em particular, no comportamento das bactérias na natureza são fenótipos complexos que resultam não só da ação de um ou de alguns genes, mas da interação de vários genes e, entre esses genes e o meio ambiente [VILAS-BÔAS et al., 2007].

Desta forma, este trabalho visa contribuir ao esclarecimento sobre a proximidade genética entre os três principais constituintes do grupo do *Bacillus cereus* sensu lato, mediante identificação e caracterização de RNAs não codificadores na espécie de *B. thuringiensis*, auxiliando na compreensão taxonômica, no entendimento do funcionamento dos fatores de virulência, na interação patógeno hospedeiro e em interações ecológicas como o comportamento de *B. thuringiensis* no controle de pragas da agricultura e vetores de doenças.

As próximas seções apresentam mais detalhes que abordam o trabalho proposto. A seção 2 apresenta a justificativa para o desenvolvimento do trabalho. As seções 3 e 4 abordam o objetivo geral e específicos propostos no presente trabalho.

2. JUSTIFICATIVA

O grupo do *B. cereus*, reúne microorganismos de grande importância que se destacam devido ao impacto significativo na saúde humana, na agricultura e na indústria de alimentos [LIU et. al., 2015]. A bactéria *B. thuringiensis* produz inclusões cristalinas protéicas, que conferem atividade entomopatogênica para várias espécies pertencentes a várias ordens de insetos, sendo utilizado no controle biológico de pragas e vetores de doenças. Por outro lado, *B. cereus* pode ser prejudicial aos seres humanos por causar intoxicação alimentar [CHAVES, 2015]. Já *B. anthracis* é conhecida como um patógeno de animais, agindo inclusive na espécie humana, sendo responsável pela doença aguda antraz e capaz de produzir uma toxina letal utilizada como arma biológica [HELGASON et. al., 2000].

Morfológicamente foi demonstrada a semelhança entre esporos e células vegetativas em *B. thuringiensis*, *B. anthracis* e *B. cereus*, porém as diferenças fenotípicas entre as três bactérias fazem com que sejam classificadas como espécies distintas [NADAL, 2015; ROSSI et. al, 2018]. Alguns genes responsáveis pelas características fenotípicas diferentes de *B. thuringiensis* e *B. anthracis* estão presentes em grandes plasmídeos [CHAVES, 2015]. Estes plasmídeos podem ser

transferidos para outras linhagens do grupo *B. cereus* através de mecanismos de transferência horizontal de genes como a conjugação e transdução, além de rearranjos genéticos promovidos por elementos móveis de DNA [VILAS-BÔAS et al., 2007].

A dificuldade de resolução taxonômica entre as linhagens de *B. cereus* e *B. thuringiensis* deve-se à intensa similaridade de seus genomas além do baixo grau de polimorfismo genético [NADAL, 2015]. A ambiguidade taxonômica do grupo ilustra a dificuldade para a definição das espécies bacterianas, logo estudos genômicos e populacionais tornaram-se relevantes para espécies que compõem o grupo *B. cereus* [CHAVES, 2015].

Para auxiliar na compreensão destas espécies, tem se mostrado relevante acrescentar aos estudos genéticos, estudos relacionados à ecologia das linhagens do grupo *B. cereus* [NADAL, 2015]. A soma destas duas áreas pode, de forma elucidativa, contribuir para a definição taxonômica deste grupo.

Assim sendo, este trabalho visa contribuir ao esclarecimento sobre a proximidade genética entre estes três constituintes do grupo do *Bacillus cereus* sensu lato mediante identificação e caracterização de RNAs não codificadores em uma espécie de *B. thuringiensis*, auxiliando na compreensão taxonômica, no entendimento dos fatores de virulência, na interação patógeno hospedeiro e em interações ecológicas como o comportamento de *B. thuringiensis* no controle de pragas da agricultura e vetores de doenças.

2.1 REVISÃO BIBLIOGRÁFICA

1.1 Características do grupo *Bacillus cereus*

O grupo *Bacillus cereus* inclui oito espécies de bactéria que estão intimamente relacionadas, sendo estas: *Bacillus mycoides*, *Bacillus pseudomycoides*, *Bacillus weihenstephanensis*, *Bacillus cytotoxicus* e *Bacillus toyonensis*, *B. cereus*, *B. anthracis*, *B. thuringiensis* [ROSSI et. al., 2018], sendo as três últimas espécies consideradas as mais importantes do grupo [VILAS-BÔAS et. al., 2007].

Alguns estudos recentes propõem que as bactérias *B. anthracis*, *B. cereus* e *B. thuringiensis* podem ser consideradas uma única espécie designada *B. cereus sensu lato*, devido à similaridade genética que existe entre elas [BATISTA et al, 2018; CHAVES, 2015; GONÇALVES, 2019]. As espécies deste grupo, compartilham um nível elevado de similaridade genética, de modo que tanto a hibridização DNA-DNA, quanto a análise de sequências do RNAr 16S e 23S, não conseguiram separar claramente os variantes de uma única espécie [CHAVES, 2015].

Apesar da similaridade genética que existe entre os constituintes do grupo *cereus*, eles apresentam características fenotípicas distintas que muitos pesquisadores acreditam estar relacionadas com a regulação gênica. Em bactérias, os ncRNAs desempenham papel importante na regulação de genes, ajustando a fisiologia bacteriana em resposta a sinais ambientais [APPEL,2018; WATERS; STORZ, 2009].

A distinção entre estas espécies do grupo *B. cereus* é atribuída essencialmente às suas características fenotípicas e às suas propriedades patogênicas, cujos genes que as codificam estão, muitas vezes, localizados em plasmídeos [CHAVES, 2015]. A Figura 1 apresenta a árvore filogenética do grupo *B. cereus*.

Nas próximas seções são apresentados mais detalhes sobre as três principais bactérias constituintes do grupo *cereus sensu lato*.

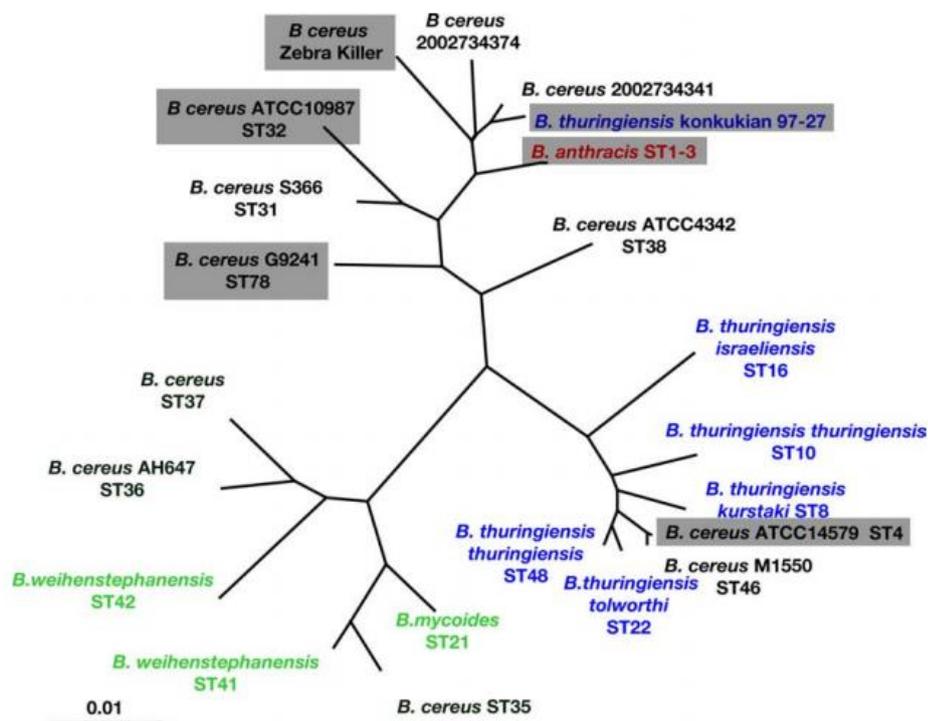


Figura 1: Árvore filogenética do grupo *Bacillus cereus* (Extraído de RASKO et al, 2006)

1.1.1 *Bacillus anthracis*

B. anthracis é um bastonete Gram-positivo aeróbico ou anaeróbico facultativo, quimiorganotrófico e imóvel [CHAVES, 2015]. Esta bactéria também se caracteriza por ser esporogênica resistente à desidratação, calor, radiação ultravioleta e até mesmo a alguns desinfetantes utilizados em serviço de saúde [CARDOSO et. al., 2015]. Os esporos desta bactéria podem permanecer viáveis no solo por muitos anos, podendo contaminar animais e humanos após contato com o agente [CARDOSO et. al, 2015].

A patogenicidade de *B. anthracis* baseia-se na presença de fatores localizados em dois plasmídeos, comumente designados como de virulência, pXO1 e pXO2 [JENSEN et. al., 2003]. Esta bactéria é responsável por causar uma doença grave conhecida como antraz, que causa severos comprometimentos dos pulmões, pele e intestino [SOUZA et al, 2004]. Devido a ação tóxica que a *B. anthracis* apresenta, sabe-se que ela tem o poder de agir como arma biológica, já tendo sido espalhada através de pó, cartas ou objetos em ações de bioterrorismo, cujo escopo é infundir histeria e pânico pela doença produzida em muitos indivíduos ao mesmo

tempo, e com isto afetar as estruturas de um país, enfraquecendo-o [RABINOVITCH & LOURENÇO, 2002].

1.1.2 *Bacillus cereus*

A espécie *B. cereus* é uma bactéria ubíqua, amplamente encontrada na natureza, sobretudo no solo. Apresenta-se em forma de bastonete, é Gram-positiva aeróbica facultativa, sendo caracterizada pela sua capacidade de formação de endósporos [CHAVES, 2015]. É uma bactéria conhecida por causar intoxicação alimentar, podendo provocar síndromes dos tipos: emética e diarréica [BATISTA et. al., 2018]. *B. cereus*, é transmitida através de alimentos contaminados. Além disso, também apresenta uma grande preocupação para a indústria de laticínios, pois possui excelente capacidade de aderir às superfícies de aço inoxidável, causando deterioração dos produtos e comprometimento dos equipamentos [BATISTA et. al., 2018].

1.1.3 *Bacillus thuringiensis*

Ao contrário de *B. anthracis* e *B. cereus*, citadas anteriormente, a bactéria *B. thuringiensis* (Bt) é a única patogênica para insetos. Ela se caracteriza por ser Gram-positiva, esporulante, aeróbica facultativa, mesófila e quimioheterotrófica, sendo que suas células vegetativas possuem a forma de bastonete [CHAVES, 2015].

B. thuringiensis é um habitante natural dos sistemas intestinais de certos insetos, provocando ou não doenças e/ou eventualmente morte, podendo ser liberada no solo e se proliferar quando as condições do ambiente (temperatura, humidade, etc) forem favoráveis [JENSEN et. al, 2003]. Esta bactéria produz um cristal parasporal, de natureza proteica que é ativo em larvas de uma variedade de ordens de insetos, permitindo-se assim o uso de produtos a base desta bactéria para controle biológico [VILAS-BÔAS et. al., 2007]. A produção destes cristais se dá durante a fase de esporulação sendo liberados no ambiente juntamente com o esporo.

Em geral os produtos a base de *B. thuringiensis* são compostos por uma mistura de cristais, esporos, poucas células vegetativas e ingredientes adicionados a

fórmula [VILAS-BÔAS et. al., 2007]. Contudo, apesar de a utilização de produtos a base deste microorganismo ser ainda baixa, tem aumentado a cada ano, correspondendo ao principal produto para controle biológico de uso comercial no Brasil (principalmente quando utilizado dentro de programas de Manejo Integrado de Pragas - MIP) e no mundo atualmente [MARCELINO, 2019; ANGELO et al, 2010].

Estima-se que os produtos a base desta bactéria correspondam a cerca de 90% do mercado mundial de bioinseticidas, sendo que a maior parte das pesquisas envolvendo Bt, bem como o maior número de trabalhos, são publicados em sua maioria em países com grandes interesses econômicos na agricultura, como é o caso dos Estados Unidos da América, China, Índia e Brasil [MARCELINO, 2019; TAMEZ-GUERRA et al, 2001].

1.1.4 Genômica de *Bacillus thuringiensis*

A sequência genômica, seja ela de qualquer organismo, apresenta informações importantes para os pesquisadores, pois permitem a descoberta de informações relevantes que ajudam a caracterizá-lo e assim compreender melhor suas relações evolutivas [APPEL, 2018; LIU et al., 2012; WENFEI et al., 2014].

De acordo com o NCBI (*National Center for Biotechnology Information* disponível em: <https://www.ncbi.nlm.nih.gov/>), os genomas das linhagens de *B. thuringiensis* possuem entre 5.343 e 7.265 genes e variam de 5,24 Mb a 6,87 Mb.

B. thuringiensis é considerada uma bactéria ecológica, fenotípica e geneticamente diversa [APPEL, 2018; SAADAOUI et al., 2010], o que levaria a uma grande variabilidade genética dentro desta espécie. As variações encontradas em *B. thuringiensis* podem ser explicadas em parte, porque esta espécie (assim como outras bactérias) pode facilmente trocar ou adquirir material genético com outras bactérias por meio de diversos eventos, sendo que alguns deles são consequência da ação de elementos genéticos móveis [MENDES, 2005]. Estes eventos promovem expansões e rearranjos genômicos que possibilitam a transferência horizontal de genes apresentando impacto notável na evolução, virulência, patogenicidade e simbiose desses organismos [APPEL, 2018; ALVARENGA et al., 2018].

B. thuringiensis também abriga uma grande variedade de elementos transponíveis, incluindo sequências de inserção (IS231, IS232, IS240, ISBt1 e ISBt2) e transposons (Tn4430 ou Tn5401), frequentemente associados aos genes *cry* (que

são os genes codificadores para a proteína Cry, responsável pelas características entomopatogênicas¹) [MAHILLON et al., 1994; LÉONARD; CHEN; MAHILLON, 1997; MISHRA, 2017; FIEDORUK, 2017; APPEL, 2018].

Acredita-se que esse “ambiente genético” (presente em plasmídeos e próximo a elementos móveis) particular dos genes *cry*, seja responsável por sua expansão em células bacterianas e por facilitar sua recombinação e troca entre plasmídeos, fornecendo uma fonte de variabilidade em linhagens produtoras de cristal gerando novas combinações e conseqüentemente novos espectros de ação [AROSON 1993 ; LÉONARD et al., 1997 ; SCHNEPF et al., 1998 ; PALMA et al., 2014].

B. thuringiensis também possui grande diversidade em seus plasmídeos [APPEL, 2018]. A maioria das linhagens de *B. thuringiensis* contém de 2 a 12 plasmídeos, com tamanhos variando entre 2 kb a 600 kb representando cerca de 10 a 20% do conteúdo genômico total desta bactéria [HAN, et al., 2006; ZHONG et al., 2011].

A relevância dos plasmídeos nas linhagens de *B. thuringiensis* é assumida não só pela possibilidade de discriminação das linhagens, mas principalmente por serem fundamentais para a patogenicidade da espécie, abrigando os genes *cry* [BERRY et al., 2002; LOEZA-LARA et al., 2005; ROH et al., 2007].

1.2 RNAs não codificantes

O DNA é responsável por conter as informações genéticas das células, informações estas que são transcritas em RNAs. O RNA por sua vez, é decodificado pelos ribossomos que converte as informações genéticas contidas nele em uma sequência protéica [ARRIAL, 2008].

As proteínas produzidas a partir do processo de tradução do RNA são as substâncias mais importantes do organismo, por desempenhar inúmeras funções e por serem fundamentais em praticamente todos os processos biológicos conhecidos [REPOILA & DARFEUILLE, 2009]. Porém, nem todas as moléculas de RNA codificam para uma proteína, sendo estes RNAs conhecidos como RNAs não codificantes (ncRNA). Um ncRNA é definido como qualquer molécula de RNA que

¹ Significa que é específica para controlar o inseto-alvo e, portanto, completamente inofensiva a qualquer ser vivo

não é traduzida em proteína, mas que está envolvida em diversos processos celulares [EDDY, 2001; COSTA, 2005; XUE; LI, 2008; SHANMUGAM; NAGARAJAN; PRAMANAYAGAM, 2017], sendo observados em organismos nos três domínios da vida: *Bacteria*, *Archaea* e *Eukarya* [FLICEK et al., 2011; KERSEY et al., 2014].

Estes RNAs são conhecidos a algum tempo, no entanto, acreditava-se que os RNAs não codificadores de proteínas não possuíam funções importantes no organismo, chegando a ser incluído em regiões chamados de “DNA lixo” [ARRIAL, 2008]. Mais tarde, após a retomada os estudos sobre os ncRNAs, alguns estudos observou que cerca de 98% do que é transcrito pelo genoma humano é constituído de ncRNAs, o que indica que uma quantidade tão grande de ncRNAs certamente deveria possuir papéis importantes no organismo [ARRUDA, 2015; BROSNAN; VOINET, 2009]. Desta forma, hoje são descritos como importantes em diversos eventos biológicos no organismo humano, incluindo a compensação de dose e inativação do cromossomo X, resposta às condições de estresse, regulação de atividades de proteínas no controle transcricional e regulação do desenvolvimento [APPEL, 2018; SUZUKI; HAYASHIZAKI, 2004].

Em bactérias, os ncRNAs desempenham papel importante na regulação de genes, principalmente a nível pós-transcricional, pois ajustam a fisiologia bacteriana em resposta a sinais ambientais; regulam a resistência a antibióticos: alteram a eficiência da tradução e/ou a estabilidade de mRNA, além de modular a atividade de proteínas [APPEL, 2018; WATERS; STORZ, 2009].

Assim sendo, genericamente existem duas classes de RNAs, os codificadores de proteínas e os não codificadores de proteínas (Figura 2). Os ncRNAs não codificam proteínas, porém participam em grande parte dos processos de regulação da expressão gênica, atuando em eventos transcricionais, pós-transcricionais e traducionais [ARRUDA, 2015]. Além disso, diferentes classes de ncRNAs foram relatadas em processos de regulação, tais como: pequenos e longos RNAs.

Os pequenos ncRNAs são caracterizados por apresentar números de bases que variam entre 20 a 30 bases [VIEIRA, 2016]. Já os longos ncRNA têm como características básica o tamanho, pois devem ser maiores do que 200 nucleotídeos, além do fato de que as proteínas não são sintetizados a partir deles [HUART et al, 2010 et al; PONTING, 2009].

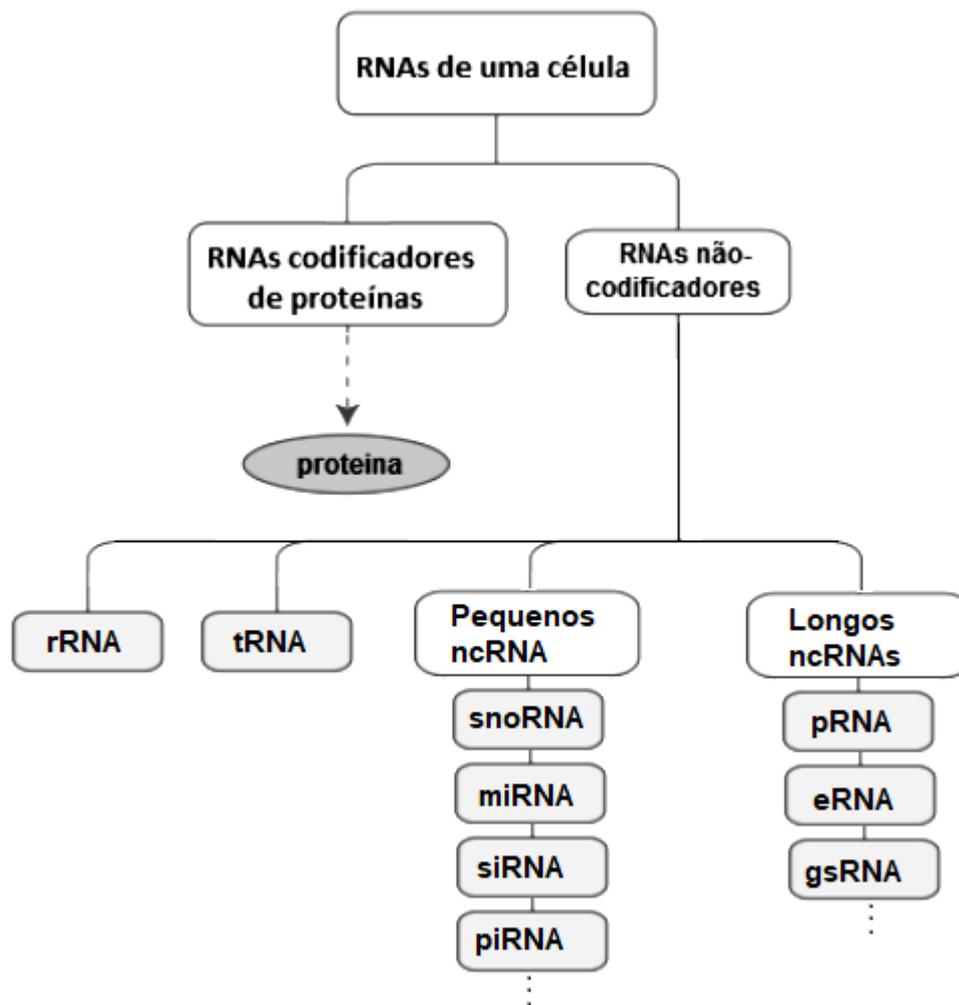


Figura 2: Duas classes de RNAs - os codificadores e não codificadores (Adaptado de KALANTIDIS, 2015)

Os RNAs não-codificantes podem ser de vários tipos e desempenhar funções variadas dentro da célula. Dentre os principais ncRNAs podemos destacar: tRNA (RNA transportador), rRNA (RNA ribossomal), snoRNA (RNA pequeno nucleolar), miRNA (micro RNA), siRNA (RNA silenciador), piRNA (Pwi-interacting) (CORREIA, 2007).

O tRNA é um dos pequenos ncRNAs mais conhecidos, sendo este responsável pelo transporte de aminoácidos que são usados na síntese de proteínas [VIEIRA, 2016]. Os tRNAs (Figura 3, estrutura secundária do tRNA) são moléculas transportadoras da informação de cada códon componente do mRNA (RNA mensageiro) em um aminoácido específico que é adicionado à proteína que esta sendo formada. O tRNA desempenha essa função através de duas regiões, o

anticódon que é responsável pelo reconhecimento de códons específicos do mRNA, e o aminoácido correspondente ao códon [PAVON-ETERNOD et al, 2009].

Outro ncRNA bastante conhecido é o rRNA (a Figura 4 mostra a estrutura secundária do rRNA), este compõe o ribossomo que tem a função de prover um mecanismo para decodificar o mRNA em aminoácidos e interagir com os tRNAs durante a tradução, na síntese de proteína [EDDY, 2001].

Já os snoRNAs (pequeno RNA nucleolar) são classes de pequenas moléculas que realizam modificações químicas no rRNA, e em outros ncRNAs, como o tRNA. O objetivo dessas modificações é promover a maturação desses ncRNAs para transformá-las em moléculas ativas [DURBIN et al, 1998].

O miRNA (microRNA) é parcialmente complementar as moléculas de mRNA, possuindo a função de reduzir a expressão de genes codificadores e inibir a tradução de mRNAs [MENDELL, 2005]. Além disso, os miRNAs podem modificar outros ncRNAs (como rRNAs) e agir como reguladores no processo de tradução, respectivamente [VIEIRA, 2016].

O piRNA (*piwi-interacting*) são existentes basicamente nas células de mamíferos e assim como os miRNAs e os snoRNAs, os piRNAs também estão relacionados com a regulação gênica, atuando no silenciamento de genes capazes de se auto duplicar no interior do genoma [ARRUDA, 2015; BRENNECKE et al, 2007].

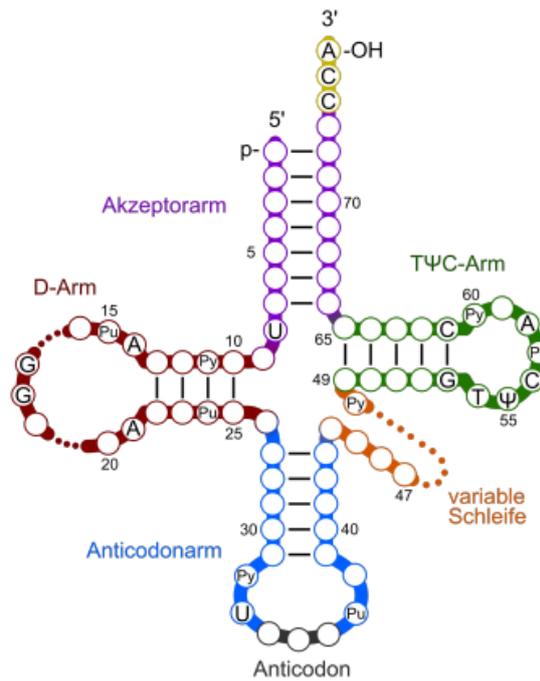


Figura 3: Estrutura secundária de um tRNA. Extraído de [VIEIRA, 2016]

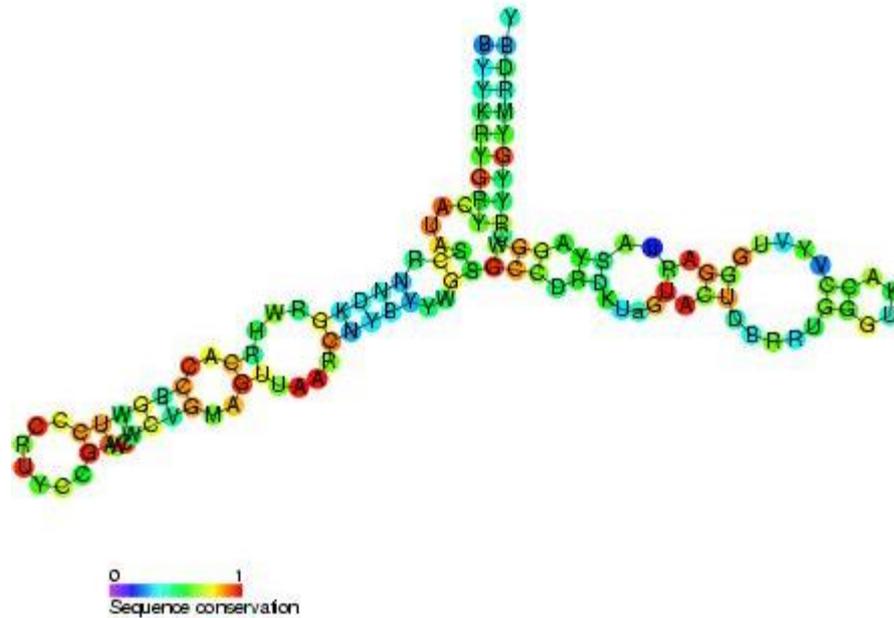


Figura 4: Estrutura secundária do rRNA. Extraído de Rfam.

Embora ainda não haja um consenso quanto a uma definição formal do que seja um ncRNA, o que se sabe, é que além de não codificarem para produtos protéicos, eles possuem uma atividade biológica relevante e não servem como intermediário entre DNA e proteína [ARRIAL, 2008]. A dificuldade de classificar ncRNAs é justificável devido a variedade de estruturas, funções e interações que

exercem *in vivo*, porém, independente de classificações, a quantidade de ncRNAs identificados cresce rapidamente na literatura [ARRIAL, 2008].

1.3 Bioinformática na busca por ncRNAs

Com o surgimento dos seqüenciadores automáticos de DNA, houve uma explosão na quantidade de sequências a serem armazenadas e que exigiam recursos computacionais cada vez mais eficientes. Além do armazenamento, também surgiu a necessidade de análise desses dados, o que tornava indispensável à utilização de plataformas computacionais eficientes para a interpretação dos resultados obtidos, nascendo assim a bioinformática [PROSDÓCIMO et. al., 2002].

O termo bioinformática refere-se a métodos computacionais para o armazenamento, processamento, análise, previsão e modelagem de dados biológicos com ajuda das ciências e tecnologias da computação [ALVES, 2013]. Além disso, ela envolve tanto engenharia de software, matemática, estatística, ciência da computação, como dados fornecidos pela biologia molecular.

Na Engenharia de Software são apresentadas duas definições que podem ser destacadas para estudos experimentais: o conceito *in vitro*, representando experimentos em que o ambiente é apresentado por meio de modelos computacionais e que descrevem a realidade com o qual os participantes podem interagir diretamente. O outro conceito, *in silico*, descreve onde tanto os participantes quanto o ambiente e o objeto de estudo são descritos como modelos computacionais, não havendo qualquer tipo de intervenção humana [TRAVASSOS & BARROS, 2003; NADAL, 2015].

A bioinformática permite diversas aplicações em genômica, que associada às tecnologias de alto desempenho na área da genética permite o uso de abordagens abrangentes no estudo da estrutura, organização de genomas, expressão diferencial, predição e classificação funcional de genes, além de demonstrar sua relação com proteínas [ALVES, 2013; APPEL, 2018].

De maneira geral pode-se considerar três categorias principais de métodos computacionais para predição de ncRNAs, sendo: (1) métodos baseados em comparações entre sequências e estruturas de RNA conservadas de forma evolutiva; (2) métodos *de novo* utilizando a sequência de RNA e características

estruturais e (3) métodos baseados no sequenciamento de transcritos seguido de montagem.

Para a predição de ncRNAs podem ser utilizadas abordagens computacionais, localizando sequências de DNA conservadas e que estão presentes em regiões intergênicas de espécies relacionadas filogeneticamente. A principal idéia é a análise de potencial de conservação das estruturas secundárias e trechos com algum grau de similaridade entre os genomas [APPEL, 2018]. Na identificação de ncRNAs, esta técnica pode ser aplicada por meio de varredura em genomas, especialmente em casos em que os organismos possuem pouca ou nenhuma informação de ncRNAs [HERTEL et al., 2009; SONG et al., 2009].

Os métodos baseados em sequência exploram a descoberta de que as frequências de muitos k -mers para ncRNAs em uma família específica geralmente compartilham distribuição de probabilidade similar, permitindo prever ncRNAs a partir desta frequência [APPEL, 2018]. Também é possível prever por meio de características estruturais, considerando atributos físico-químicos como a estabilidade termodinâmica da molécula.

Atualmente, as diferentes abordagens permitem identificar um grande número de novos ncRNAs, reafirmando a importância da análise computacional para atender indagações relacionadas ao número e diversidade de genes de RNAs não codificantes [APPEL, 2018].

A genômica comparativa é considerada uma das abordagens mais modernas na área do sequenciamento total de genomas. Esta consiste na análise e comparação do material genético de diferentes espécies, com o propósito de estudar a estrutura, organização e evolução dos genomas, juntamente a funções dos genes e regiões não codificantes [CATANBO; DEGRAVE; DE MIRANDA, 2007]. Com base na comparação de dois organismos semelhantes pode-se utilizar a similaridade observada para identificar a localização dos genes do segundo indivíduo com base no primeiro [SEPÚLVEDA, 2015].

Outra grande utilidade da genômica comparativa é a construção de árvores filogenéticas que são representadas por um grafo na forma de uma árvore, que explica o processo evolucionário e é construída a partir da sequência de atributos (tais como sequências de DNAs) obtidos de diversos organismos [MATSUDA et al, 1999]. As diferenças são muito específicas em organismos com um ancestral

próximo. Pode-se seguir o percurso filogenético de um gene e organizar os seres em classes segundo o número de genes em comum [PRADO, 2001; SEPÚLVEDA, 2015].

Este tipo de abordagem está ajudando cada vez mais os pesquisadores a elucidarem as adaptações moleculares em genes e proteínas, necessárias para que um organismo consiga ou não viver em determinado ambiente ou a metabolizar compostos que apresentam alguma característica direta ou indiretamente relacionada à política financeira ou ambiental, síntese de antibióticos ou degradação de xenobióticos respectivamente [PROSDÓCIMI & MOREIRA,]. Além disso, quando relacionada com a evolução permite o desenvolvimento de vários campos do conhecimento, destacando: analisar a origem e propagação de doenças e os seus vetores; realizar bioprospecção para produtos farmacêuticos e agrícolas como produtos para o controle biológico indicado para espécies invasoras além de examinar fatores de risco para a conservação de espécies e restauração dos ecossistemas [SEPÚLVEDA, 2015].

3. OBJETIVO GERAL

O objetivo geral deste trabalho, visa a caracterização dos RNAs não codificadores descritos para *Bacillus thuringiensis*, a partir de bancos de dados de sequências enriquecidas com RNAs não codificantes na linhagem Bt407.

3.1 Objetivos específicos

Os objetivos específicos propostos por este trabalho são descritos a seguir:

- Utilizar técnicas de bioinformática para a identificação de RNAs não codificadores no genoma expresso enriquecido com sequências de ncRNAs de *B. thuringiensis*;
- Fornecer informações para possibilitar uma melhor compreensão do papel dos RNAs não codificadores na regulação dos genes em processos biológicos e fatores de virulência.

- Comparar as estratégias de busca direta no genoma com anotação de sequências expressas enriquecidas com ncRNAs.

4. MATERIAIS E MÉTODOS

4.1 Ferramentas computacionais

A Tabela 1 apresenta as principais ferramentas que foram utilizados no desenvolvimento do projeto.

Tabela 1: Ferramentas utilizadas no projeto

Ferramenta	Função
Infernal	Pesquisa de estruturas de RNA a partir de sequências de bancos de dados de DNA
Rfam	Classificação de famílias de ncRNA
Artemis	Visualização de sequências e extração de regiões intergênicas
Bowtie	Alinhamento de sequências
RNAfold	Prever estruturas secundárias de sequências de DNA ou RNA de fita simples

Nas próximas seções serão descritos mais detalhes sobre as ferramentas utilizadas e a finalidade para estas estabelecidas.

4.1.1 Infernal/ Rfam

O Infernal (*INFERENCE of RNA ALIGNMENT*) é um pacote de software que permite criar perfis de estrutura secundária de RNA consenso e usá-los para pesquisar bancos de dados de sequências para RNAs homólogos, ou para criar novos alinhamentos de sequência múltipla [NAWROCKI & EDDY, 2016]. Esta ferramenta utiliza um caso especial de gramáticas livre de contexto estocástico, chamados modelo de covariância (CM).

Um CM assemelha-se a um perfil de sequência, porém, pontua uma combinação entre o consenso de sequência e o consenso de estrutura secundária do RNA. Desta forma, é mais provável identificar RNAs homólogos que conservam mais sua estrutura secundária do que sua estrutura primária [NADAL, 2015].

O pacote INFERNAL pode ser usado com o Rfam para anotar seqüências (incluindo genomas completos) para homólogos a ncRNAs conhecidos.

O Rfam é um banco de dados que contém informações de famílias de ncRNAs e outros elementos de RNAs estruturados. É um banco de dados de acesso livre desenvolvido pelo *Wellcome Trust Sanger Institute* em colaboração com o instituto *Janelia Farm* (O Rfam esta disponível em: <https://rfam.xfam.org/>).

O Rfam divide os ncRNAs em famílias baseadas na evolução de um ancestral comum. A produção de alinhamentos de múltiplas seqüências dessas famílias pode fornecer informações sobre sua estrutura e função.

No site do Rfam é possível pesquisar ncRNAs por palavra-chave, nome de família ou genoma, assim como pesquisar por uma seqüência de ncRNA ou número de acesso EMBL. As informações do banco de dados também estão disponíveis para download, instalação e uso através do pacote de software INFERNAL.

4.1.2 Artemis

O Artemis é um software utilizado para visualizar sequencias de DNA e também como ferramenta de anotação gênica, que permite a visualização da seqüência, de maneira esquemática e os resultados das análises dentro do contexto da seqüência e sua tradução de todos os seis polipeptídeos possíveis. O Artemis é escrito em Java e lê seqüências de formato EMBL ou GENBANK e tabelas de recursos, podendo trabalhar com seqüências de qualquer tamanho [CARVER et al, 2008].

4.1.3 Bowtie

O Bowtie é um pacote de software usado para alinhamento e análise de seqüências em bioinformática e distribuído livremente para Windows, Linux e

MacOS. O Bowtie é um alinhador ultrarrápido e eficiente na memória para sequências curtas de DNA, normalmente utilizado com leituras curtas oriundas de Sequenciamento de Nova Geração (NGS) e utiliza um longo genoma como referência, ou para análise genômica completa [LANGMEAD et al, 2019].

4.1.4 RNAfold

O programa RNAfold analisa as moléculas do ncRNA e é um preditor de estrutura secundária através da probabilidade de que as regiões de ligação estejam acessíveis em suas estruturas secundárias prevendo interações intra-moleculares [BERNHART; HOFACKER; STADLER, 2005].

4.2 Estratégias para busca e caracterização de ncRNAs

4.2.1 Mineração de dados

Nesta etapa a sequência completa do genoma de *B. thuringiensis* foi obtida por meio de busca no banco de dados público NCBI (*National Center for Biotechnology Information*). A linhagem utilizada neste trabalho foi a Bt407 que é apresentada na Tabela 2 juntamente com os respectivos códigos de acesso às sequências de seu genoma.

Tabela 2: Linhagem Bt 407 e as identificações e os respectivos códigos de acesso das sequências do genoma

Linhagem Bt 407ASM30674v1
Cromossomo CP003889.1
Plasmídeo BTB_15p CP003892.1
Plasmídeo BTB_2p CP003897.1
Plasmídeo BTB_502p CP003890.1

**Plasmídeo BTB_5p
CP003896.1**

**Plasmídeo BTB_6p
CP003895.1**

**Plasmídeo BTB_78p
CP003891.1**

**Plasmídeo BTB_7p
CP003894.1**

**Plasmídeo BTB_8p
CP003893.1**

**Plasmídeo BTB_9p
CP003898.1**

O acesso e download do genoma de referência de *B. thuringiensis* utilizado neste estudo, foi realizado no dia 28 de novembro de 2018.

4.2.2 Extração de sequências de RNA

Nesta etapa foram utilizados dados obtidos através da extração e seqüenciamento de RNA total de *B. thuringiensis* pelo grupo de pesquisas da UEL. Assim, a linhagem Bt407 foi cultivada a 37 °C com 140 rpm de agitação orbital até entrada em fase estacionária. Em seguida, amostras de RNA total da linhagem foi extraída nos tempos T-0 (final de fase logarítmica de crescimento) e T-1 (início de fase estacionária) utilizando o kit Illustra™ RNAspin Mini RNA (GE Healthcare illustra™).

As amostras passaram por três etapas de purificação e centrifugação. Posteriormente, foi realizada a quantificação e análise da qualidade e integridade do RNA. O sequenciamento das amostras realizado foi enriquecido com sequências de ncRNAs. A Figura 5 faz a representação da curva de crescimento de uma bactéria, demonstrando os dois momentos (T-0 e T-1) de extração da amostra de RNA utilizadas por Appel 2018.

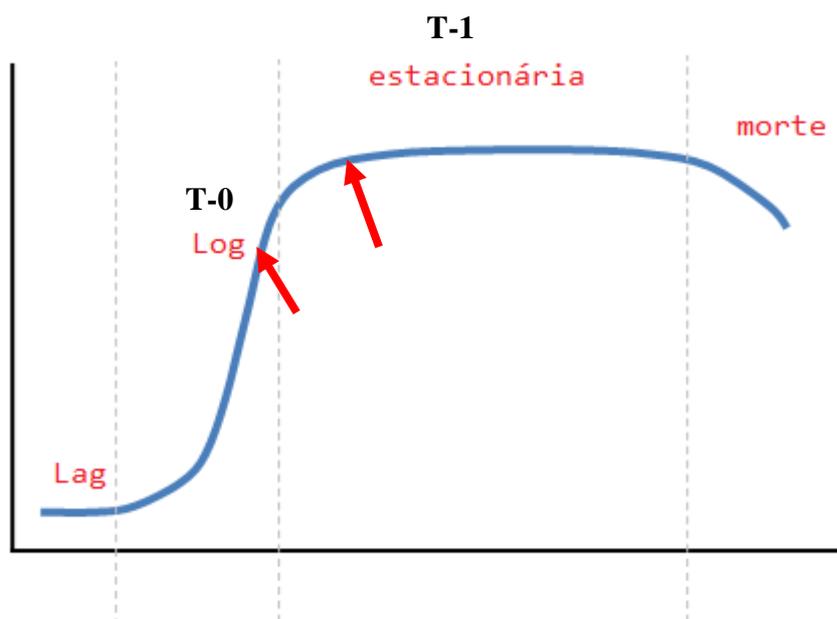


Figura 5: Gráfico representando a curva de crescimento de uma bactéria. (T-0) Final da fase logarítmica de crescimento, (T-1) Início de fase estacionária. (Autoria própria).

As amostras obtidas representam dois momentos na fase de crescimento de *B. thuringiensis* e, desta forma, pressupõe-se a possibilidade de que diferentes ncRNAs são expressos durante os dois momentos. Assim, a análise das amostras extraídas em diferentes fases de crescimento, permite avaliar quais ncRNAs estão envolvidos em cada fase ou a frequência em que são expressos em cada uma delas. As amostras extraídas em T-0 e T-1 são definidas neste trabalho como amostra 1 e amostra 2 respectivamente.

O genoma de referência de *B. thuringiensis* obtido na fase de mineração de dados”, foi utilizado no Artemis v-16.0.0 juntamente com as amostras (amostras 1 e amostra 2) mencionadas anteriormente. As amostras sequenciadas foram alinhadas utilizando a ferramenta Bowtie v-1.2.2 e, os arquivos de alinhamento obtidos foram utilizados junto do genoma de referência no visualizador no software Artemis v-16.0.0. Nesta etapa foi realizada a visualização e extração das sequências de RNA nas regiões com alta expressão.

A extração das sequências foi realizada de forma manual e, as sequências obtidas para cada uma das amostras, foram convertidas em dois arquivos multiFASTA, para serem utilizadas posteriormente na fase de anotação. A Figura 6 mostra um exemplo de região extraída para análise de ncRNAs, onde a região indicada pela seta é altamente expressa.

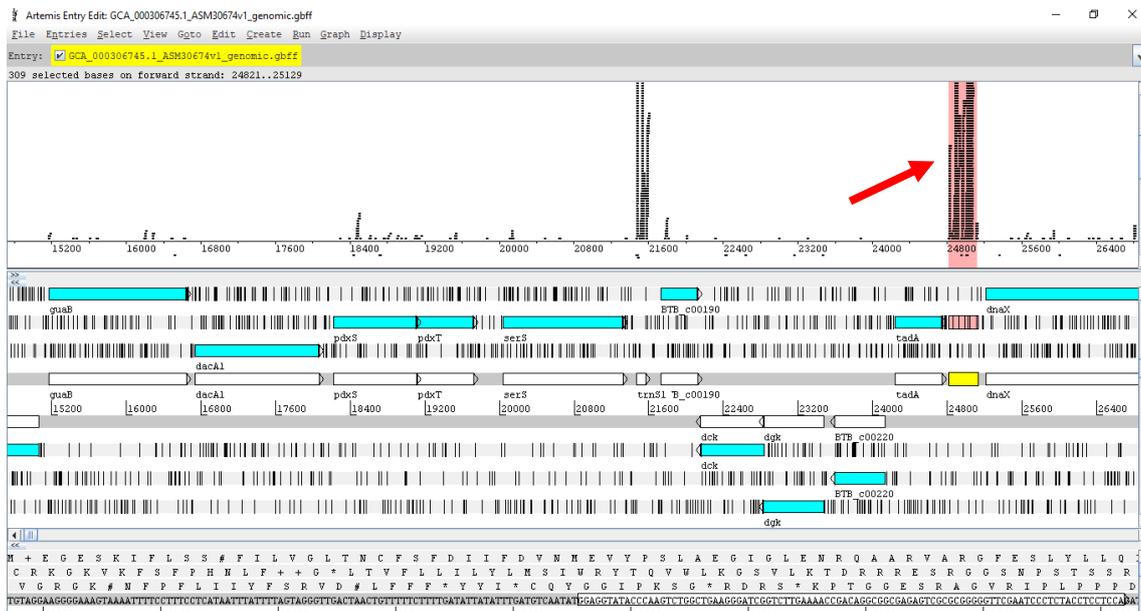


Figura 6: Visualizador Software Artemis - exemplo de região selecionada para análise de ncRNAs

4.2.3 Anotação

A anotação dos ncRNAs foi realizada utilizando a ferramenta *cmsearch* do pacote INFERNAL v-1.1.2 juntamente com os modelos de covariância (a partir de dados de *Bacillales*) do banco de dados Rfam v-13.0. Foi utilizado como parâmetros na busca dois arquivos multi fasta referentes a amostra 1 e amostra 2 (tempo 0 e tempo 1 respectivamente). Para os resultados obtidos pela ferramenta serem considerados como significativos, foi utilizado como valor padrão os limites de inclusão *E-value* de 0,01. Os resultados com valores que ultrapassaram o valor padrão, foram considerados não significativos.

Levando em consideração que o *cmsearch* pode excluir sequências significativas para o estudo devido aos valores de inclusão utilizados, uma análise posterior foi realizada manualmente, a fim de resgatar candidatos dados como não significativos. Nesta nova análise, as sequencias que possuem valor de score maior que 17.0 foram inclusos nos resultados (Tabelas 9 e 10 nos Apêndices).

4.4 Fluxograma do projeto

A Figura 7 apresenta o diagrama com a representação das atividades e ferramentas utilizadas no desenvolvimento do trabalho.

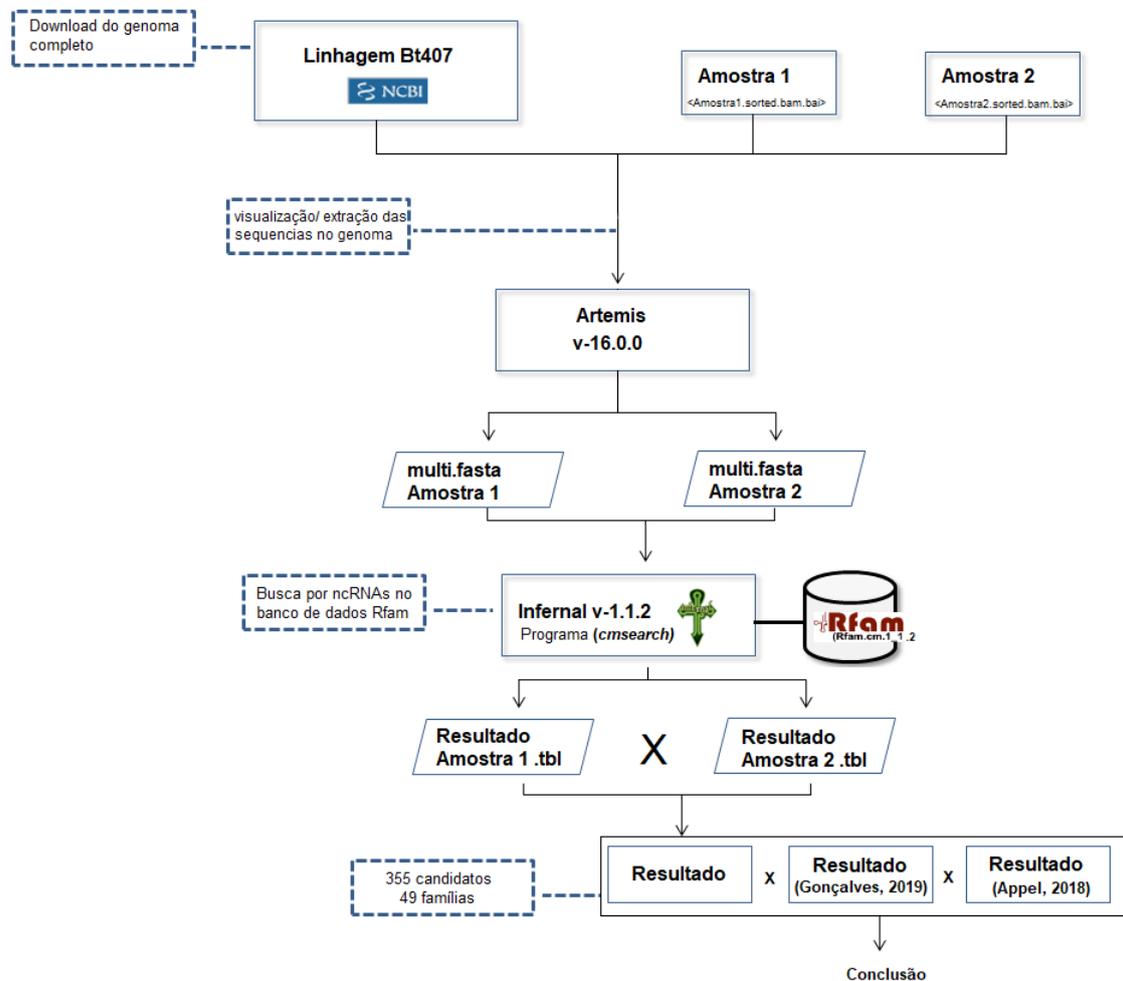


Figura 7: Fluxograma das etapas do projeto

5. RESULTADOS E DISCUSSÕES

5.1 Predição e avaliação dos ncRNAs identificados

Na primeira análise das sequências extraídas através do programa *cmsearch* do pacote INFERNAL no conjunto de sequências de *B. thuringiensis*, foram encontrados um total de 894 possíveis moléculas de ncRNAs, considerando 308

significativos (limites de inclusão máximos de 0,01 de *e-value*); sendo 131 na amostra 1 (T-0) e 177 na amostra 2 (T-1).

Dentro das 308 seqüências foram identificadas um total 49 famílias de ncRNAs diferentes. Destas 17 não eram comuns entre as duas amostras. Ademais 7 famílias encontradas na amostra 1 não eram encontradas na amostra 2 e 10 famílias exclusivas na amostra 2. A tabela 3 apresenta as famílias descritas como significativas e o total de seqüências encontradas em cada amostra.

Em uma segunda análise, foi realizada uma nova averiguação nos resultados definidos anteriormente como não significativos pelos limites de inclusão do *cmsearch*. A nova análise dos dados levou em consideração novos limites de inclusão, tomando como base o fato de que alguns candidatos são excluídos dos resultados significativos, porém estão próximos do limite de inclusão (*E-value* 0,01). Com isso, foi tomado como hipótese que resultados próximos dos valores de inclusão do *cmsearch*, podem ser relevantes para o estudo, uma vez que podem apresentar informações de ncRNAs previamente não estabelecidas.

Após a nova análise dos resultados não significativos, foram encontradas 21 candidatos na amostra 1 e 25 novos candidatos na amostra 2. Estes candidatos estão inclusos em 17 novas famílias, não presentes na estratégia anterior (Famílias marcadas com (*) na Tabela 3) e realizada uma distribuição dos novos candidatos nas famílias encontradas (tabelas 8 e 9).

Tabela 3: Relação das famílias de ncRNAs encontrados. Novas famílias atribuídas aos resultados após análise posterior dos limites de inclusão apresentam-se marcadas com (*).

Familia	Rfam	Número na Amostra 1	Número na Amostra 2
5_8S_rRNA	RF00002	1	5
5S_rRNA	RF00001	5	10
6S	RF00013	2	2
Bacteria_large_SRP	RF01854	2	1
Bacteria_small_SRP	RF00169	2	1
BsrF	RF01411	1	1
c-di-GMP-I	RF01051	1	1
FMN	RF00050	1	2
FsrA	RF02273	0	1
Glycine	RF00504	1	1
L10_leader	RF00557	4	3
L20_leader	RF00558	1	0
LSU_rRNA_archaea	RF02540	1	5

LSU_rRNA_bacteria	RF02541	1	5
LSU_rRNA_eukarya	RF02543	1	5
ncr1175	RF02450	3	4
ncr1241	RF02451	2	0
PreQ1	RF00522	0	1
PyrR	RF00515	5	6
rli33	RF01469	1	0
rli61	RF01485	0	1
rliE	RF01459	2	2
RNaseP_bact_a	RF00010	1	1
RNaseP_bact_b	RF00011	1	1
RsaJ	RF01822	1	0
RsaE	RF01820	1	3
SAM	RF00162	1	1
sau-27	RF02387	1	0
sau-5971	RF02397	2	1
SSU_rRNA_bacteria	RF00177	2	7
T-box	RF00230	5	5
tmRNA	RF00023	6	9
tRNA	RF00005	45	51
TPP	RF00059	1	3
SSU_rRNA_eukarya	RF01960	2	6
*rli52	RF01480	0	1
rli38	RF01470	0	1
RsaA	RF01816	1	0
tRNA-Sec	RF01852	37	47
*ncr1015	RF02449	1	2
*ncr982	RF02454	0	1
*Spy392987	RF02592	1	0
*SpF41_sRNA	RF02435	1	2
*tsr26	RF02579	1	0
*tsr31	RF02580	1	1
*tsr33	RF02582	1	0
*psm_mec_RNA	RF02693	1	0
*tracrRNA	RF02348	0	1
*PyrG_leader	RF02371	2	1
SOMA		153	202

Com a análise e a redistribuição dos candidatos para as duas amostras, foi possível verificar que algumas famílias de sRNAs não são comuns entre as duas amostras. A Tabela 4, apresenta as famílias não comuns entre as amostras e a quantidade de sequências significativas encontradas para cada uma delas.

Tabela 4: Famílias de ncRNAs exclusiva de cada amostra analisada

Familia	Rfam	Número da Amostra 1	Número da Amostra 2
5_8S_rRNA	RF00002	1	-
L20_leader	RF00558	1	-
ncr1241	RF02451	1	-
rli33	RF01469	1	-
RsaJ	RF01822	1	-
sau-27	RF02387	1	-
sau-5971	RF02397	1	-
FsrA	RF02273	-	1
PreQ1	RF00522	-	1
rli61	RF01485	-	1
RsaE	RF01820	-	1
TPP	RF00059	-	3
*rli52	RF01480	-	1
*ncr1015	RF02449	-	1
*tsr31	RF02580	-	1
*tracrRNA	RF02348	-	1
*ncr1575	RF02452	-	1
SOMA		7	10

Na Figura 8 é apresentado o gráfico das famílias que aparecem com maior frequência entre os candidatos. As famílias tRNA e tRNA-Sec, são as famílias que possuem maior número de candidatos expressos e também o maior número de candidatos em comum entre as amostras analisadas.

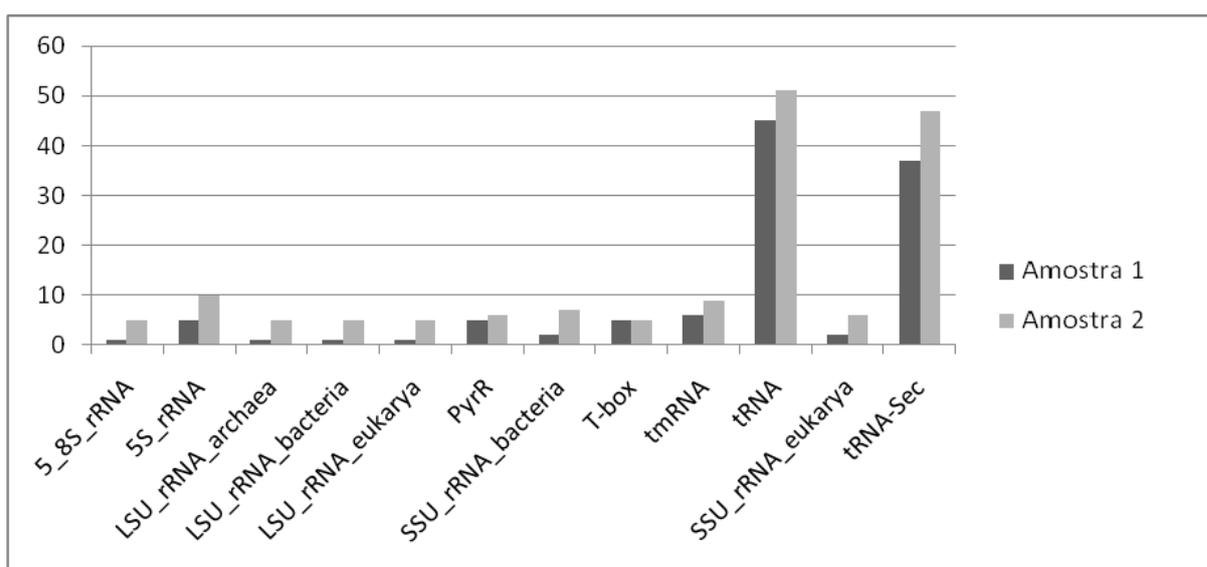


Figura 8: Gráfico da relação de candidatos para as famílias mais freqüentes nas amostras

5.2 Famílias de ncRNAs com maior número de candidatos

Algumas famílias identificadas durante o estudo apresentaram maior quantidade de candidatos em relação a outras famílias. A seguir, são descritos alguns destes ncRNAs que foram observados com maior frequência durante as amostras analisadas bem como provável função descritas nos bancos de dados.

O RNA ribossomal 5_8S_rRNA (5.8S rRNA) é um componente de RNA não codificante da subunidade grande do ribossomo eucariótico e, portanto, desempenha um papel importante na tradução de proteínas. Este é transcrito pela RNA polimerase I como parte do precursor 45S que também contém 18S e 28S rRNA [ABOU & NAZAR, 1997].

A estrutura secundária obtida através de uma sequência que representa essa família pode ser observada na Figura 9.

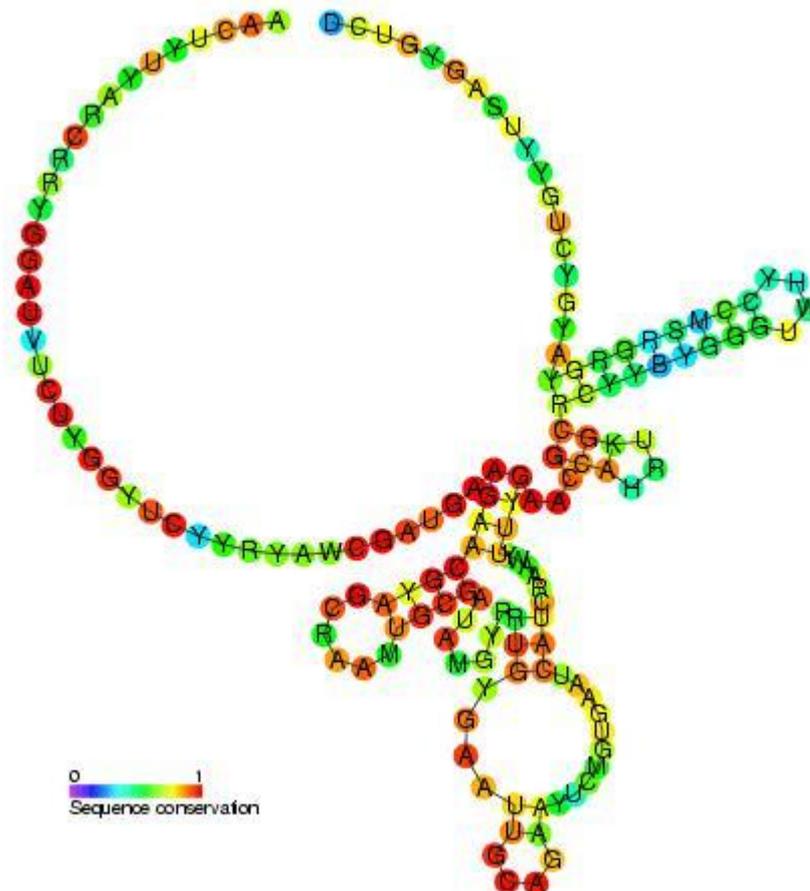


Figura 9: Estrutura secundária proposta para 5_8S_rRNA

O ácido ribonucleico ribossômico (rRNA) é o componente de RNA do ribossomo, essencial para a síntese de proteínas em todos os

organismos vivos. O rRNA é o RNA predominante na maioria das células, compondo cerca de 80% do RNA celular. Os ribossomos são aproximadamente 60% de rRNA e 40% de proteína em peso. Um ribossomo contém duas subunidades, a subunidade ribossômica grande (LSU) e a subunidade ribossômica pequena (SSU) [LAFONTAINE & TOLLERVEY, 2011].

Os ribossomos procarióticos contêm três rRNAs, que são os 23S e 5S rRNAs na LSU e o 16S rRNAs na SSU. O ribossomo procariótico contém cerca de 50 proteínas ribossômicas [LAFONTAINE & TOLLERVEY, 2011].

Os ribossomos eucarióticos e os rRNAs são maiores e mais polimórficos do que os dos procariontes. No fermento, a LSU contém os rRNAs 5S, 5.8S e 28S. Os 5.8S e 28S combinados são aproximadamente equivalentes ao rRNA 23S procariótico, exceto os segmentos de expansão (ESs) localizados na superfície do ribossomo. A SSU contém o rRNA 18S, que também contém ESs. Os ESs da SSU são geralmente menores que os ESs da LSU [LAFONTAINE & TOLLERVEY, 2011].

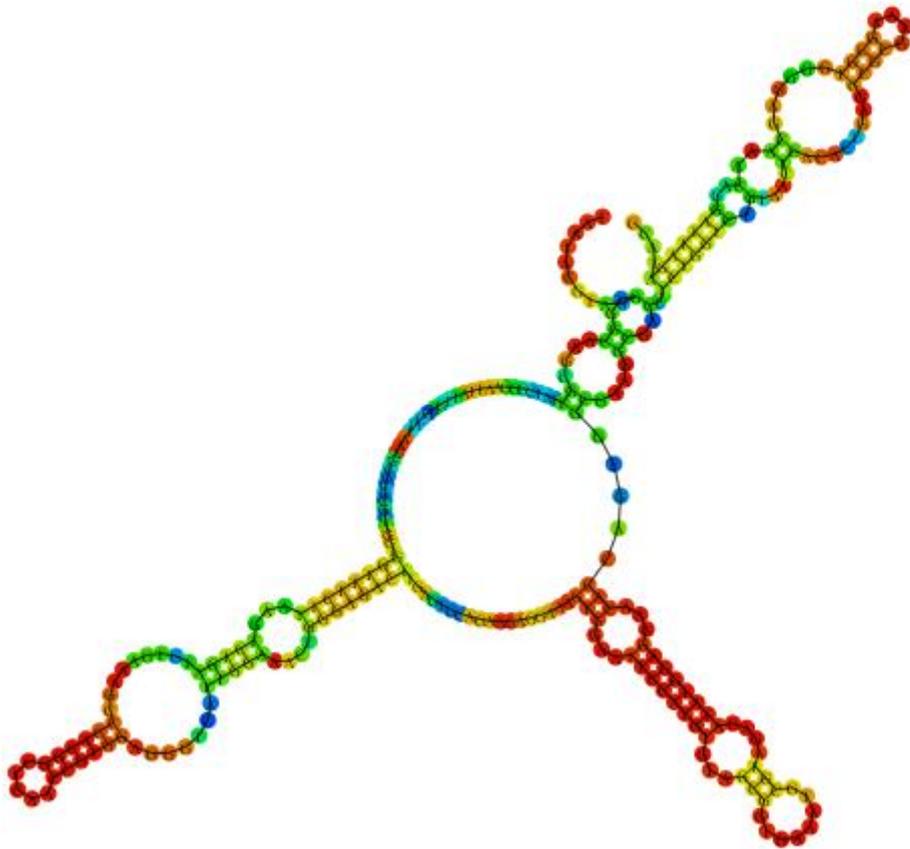


Figura 10: Estrutura secundária SSU_rRNA_bacteria

O rRNA de LSU foi chamado de ribozima, porque a proteína ribossômica não está presente no local catalítico do ribossomo (o centro da peptidil transferase, PTC). No entanto, o rRNA não demonstrou ser catalítico na ausência de proteínas. O rRNA da SSU participa da decodificação do mRNA no sítio catalítico do ribossomo. [BONNER, 2001 et al].

As sequências de SSR e LSU rRNA são também amplamente utilizadas para estudos das relações evolutivas entre organismos, uma vez que são conservadas nos grupos e são encontradas em todas as formas de vida conhecidas e sendo resistentes à transferência horizontal de genes [BONNER, 2001 et al].

O local de ligação do PyrR é um elemento de RNA que é encontrado a montante de uma variedade de genes envolvidos na biossíntese e transporte da pirimidina.



Figura 11: Estrutura secundária de PyrR

A estrutura do RNA da família PyrR (Figura 11) permite a ligação da proteína PyrR que regula a biossíntese de pirimidina em *Bacillus subtilis*. Quando a proteína se liga, forma-se um grampo finalizador a jusante, reprimindo a transcrição dos genes da biossíntese [BONNER, 2001 et al].

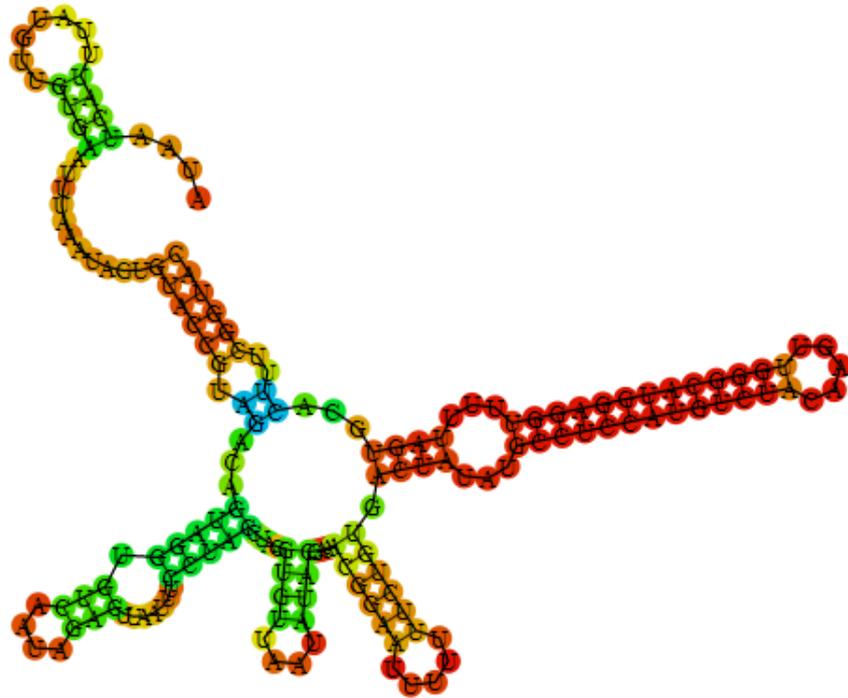


Figura 12: Estrutura secundária de L10_leader

A família L10_leader (Figura 12) é um putativo líder da proteína ribossômica de estrutura auto reguladora encontrada em *B. subtilis* e outras bactérias Gram-positivas de baixo conteúdo de GC. Está localizado nas regiões 5' não traduzidas de mRNAs que codificam as proteínas ribossômicas L20 e L12 (rplJ-rplL). Uma estrutura de terminação de transcrição independente de Rho que, provavelmente, está envolvida na regulação está incluída na extremidade 3'. Outros líderes de proteína ribossomal identificados no mesmo estudo incluem os de L13, L19, L20 e L21 [ZENGUEL & LINDAHL, 1994]

5.3 Comparação de metodologias

5.3.1 Metodologia proposta por Appel, 2018 . Predição, análise e validação de RNAs não codificadores em *Bacillus thuringiensis*

O trabalho proposto por Appel (2018) aborda a identificação, análise e validação de ncRNAs em três linhagens da espécie *B. thuringiensis*. As sequências completas de cromossomos e plasmídeos dos genomas das linhagens Bt407, HD-1 e HD73 foram selecionados. Foi realizada uma anotação gênica utilizando o programa *cmsearch* do pacote INFERNAL contra modelos de covariância do banco de dados Rfam. Os limites de inclusão utilizados como padrão para considerar o ncRNA como significativo foi de um *E-value* 0,01.

Para a distribuição dos ncRNAs nos genomas analisados, foram extraídas as coordenadas dos ncRNAs do arquivo de saída do INFERNAL e classificados em três classes: núcleo; exclusivas; e compartilhadas.

Ao todo, foram identificados 27.111 possíveis ncRNAs; sendo 1.874 candidatos considerados significativos, distribuídos em 177 famílias distintas; encontradas principalmente em regiões presentes nas três linhagens. Das famílias, 624 foram encontradas na linhagem Bt407, 656 em HD-1 e 594 em HD73.

As moléculas identificadas estão envolvidas em processos regulatórios complexos; principalmente relacionados à virulência, patogenicidade, resistência e sobrevivência da bactéria. A família epsC teve sua expressão validada no genoma.

Os resultados alcançados evidenciam que mesmo bactérias muito relacionadas filogeneticamente podem apresentar diferentes respostas às modificações ambientais, apresentando diferentes padrões de expressão; e ainda levam a uma nova linha de investigação para a compreensão de mecanismos fisiológicos em *B. thuringiensis*.

5.3.2 Metodologia proposta por Gonçalves, 2019

No trabalho de Brumatti (2019), foram extraídos das bases de dados GOLD e NCBI 132 genomas, sendo 40 de *Bacillus thuringiensis*, 44 de *Bacillus anthracis* e 48 de *Bacillus cereus*. Para anotação gênica dos genomas obtidos foi utilizada a ferramenta *cmsearch* do pacote de ferramentas INFERNAL juntamente com o modelo de covariância *Bacillales* do banco de dados Rfam.

Para *B. thuringiensis* o INFERNAL obteve um total de 1.942 moléculas de ncRNAs, onde apenas 332 foram considerados significativos. Os resultados significativos obtidos no INFERNAL, passaram por uma filtragem, onde foram removidas as famílias sobrepostas e famílias de ncRNAs comuns (tRNA, tmRNA, 58s_RNA e 5s_RNA), totalizando no final 181 moléculas distribuídas em 46 famílias de ncRNAs para *B. thuringiensis*.

5.3.3 Comparando os resultados

Nesta seção são comparados os resultados das metodologias apresentadas com a metodologia proposta neste trabalho. A Tabela 5, apresenta a comparação.

Tabela 5: Comparação dos resultados com outras metodologias

Família	Appel (2018)	Expresso	Gonçalves (2019)
U12	1	0	0
RNaseP_bact_a	1	2	0
RNaseP_bact_b	1	2	1
U3	2	0	0
6S	2	4	2
Metazoa_SRP	1	0	0
let-7	1	0	0
Intron_gpl	2	0	1
Intron_gpII	4	0	4
Histone3	1	0	1
FMN	2	3	2
TPP	8	4	8
yybP-ykoY	2	0	1
CsrC	3	0	0
S15	1	0	0
IS128	1	0	0
SAM	18	2	18
Purine	7	0	7
Lysine	4	0	4
Bacteria_small_SRP	1	3	0
Cobalamin	1	0	1
SSU_rRNA_bacteria	14	9	14
T-box	45	10	45
glmS	1	0	1
ctRNA_pND324	1	0	0
snoZ221_snoR21b	2	0	0
snoZ165	1	0	0

sroD	1	0	0
ydaO-yuaA	4	0	4
ykoK	2	0	2
IBV_D-RNA	1	0	0
Glycine	1	2	1
Thr_leader	13	0	0
Leu_leader	6	0	0
His_leader	1	0	0
PyrR	4	11	3
ylbH	3	0	2
PreQ1	2	1	2
L13_leader	1	0	0
L19_leader	1	0	1
L10_leader	1	7	1
L20_leader	2	1	1
L21_leader	2	0	2
SNORD69	1	0	0
MIR159	1	0	0
mir-147	1	0	0
mir-202	1	0	0
mir-451	2	0	0
mir-308	1	0	0
mir-14	2	0	0
mir-299	3	0	0
mir-335	1	0	0
lsy-6	1	0	0
mir-374	2	0	0
mir-259	1	0	0
MIR1122	2	0	0
MIR529	1	0	0
MIR1444	1	0	0
mir-556	6	0	0
mir-578	1	0	0
mir-583	1	0	0
mir-553	2	0	0
mir-1255	1	0	0
MIR530	1	0	0
c-di-GMP-I	4	2	4
PK-G12rRNA	14	0	0
snR77	1	0	0
BsrC	2	0	2
BsrF	1	2	1
miR-430	3	0	0
S_pombe_snR42	1	0	0
rli40	2	0	1
AdoCbl_riboswitch	1	0	0

rli61	1	1	0
rliD	1	0	1
plasmodium_snoR31	1	0	0
Bacillus-plasmid	4	0	3
SAM-I-IV-variant	1	0	0
crcB	2	0	2
epsC	1	0	1
pan	1	0	1
yjdB	1	0	1
cspA	6	0	6
greA	1	0	0
rimP	1	0	0
MicX	4	0	0
RsaE	1	4	1
RsaJ	1	1	0
tRNA-Sec	33	84	0
Bacteria_large_SRP	1	3	1
Archaea_SRP	1	0	0
Phe_leader	3	0	0
SSU_rRNA_archaea	14	0	0
SSU_rRNA_eukarya	14	8	0
group-II-D1D4-1	4	0	0
sraA	1	0	0
STnc40	1	0	0
STnc370	2	0	0
STnc320	2	0	0
STnc540	1	0	0
rivX	1	0	0
FsrA	1	1	0
TtnuCD19	1	0	0
GlsR16_miR3	1	0	0
psRNA14	2	0	0
PyrG_leader	2	3	0
SR1	3	0	2
SurA	2	0	1
sau-19	1	0	0
sau-50	3	0	0
sau-5971	5	3	0
NsiR1	1	0	0
SpF25_sRNA	2	0	0
SpF43_sRNA	1	0	0
ncr1015	2	3	0
ncr1241	1	2	1
ncr1575	2	0	2
ohsC_RNA	1	0	0
Atu_C9	4	0	0

Toxl	2	0	1
mir-965	2	0	0
LSU_rRNA_archea	14	6	0
LSU_rRNA_bacteria	14	6	14
SSU_rRNA_microsporidia	14	0	0
LSU_rRNA_eukarya	14	6	0
mt-tmRNA	1	0	0
Spy1186876	1	0	0
Spy392987	1	1	0
BSnc120	7	0	0
ncr1175	0	7	0
rli33	0	1	0
rliE	0	4	0
sau-27	0	1	0
rli52	0	1	0
rli38	0	1	0
RsaA	0	1	0
ncr982	0	1	0
SpF41_sRNA	0	3	0
tsr26	0	1	0
tsr31	0	2	0
tsr33	0	1	0
psm_mec_RNA	0	1	0
tracrRNA	0	1	0
BtsR1	0	0	1
group-II-D1D4-5	0	0	4
Soma	445	223	181

Para a comparação apresentada na Tabela 5, foram removidas as famílias de ncRNAs consideradas comuns (tRNA, tmRNA, 58s_RNA e 5s_RNA) dos resultados. Além disso, foram utilizados apenas os resultados para linhagem Bt407. Desta forma, Appel (2018) identificou para linhagem Bt407 de *B. thuringiensis* 445 moléculas ncRNAs distribuídas em 129 famílias. Na metodologia proposta por Gonçalves (2019), foram identificadas 181 moléculas de ncRNAs distribuídas em 46 famílias. A metodologia proposta aqui neste trabalho identificou (após a remoção das famílias comuns) 223 moléculas de ncRNAs distribuídas em 45 famílias. Juntas as metodologias obtiveram 849 moléculas.

Analisando os dados observa-se que 71 famílias são exclusivas de Appel (2018), 2 exclusivas de Gonçalves (2019) e 14 exclusivas da metodologia proposta. Na Figura 13, é apresentada uma comparação por meio do diagrama de venn.

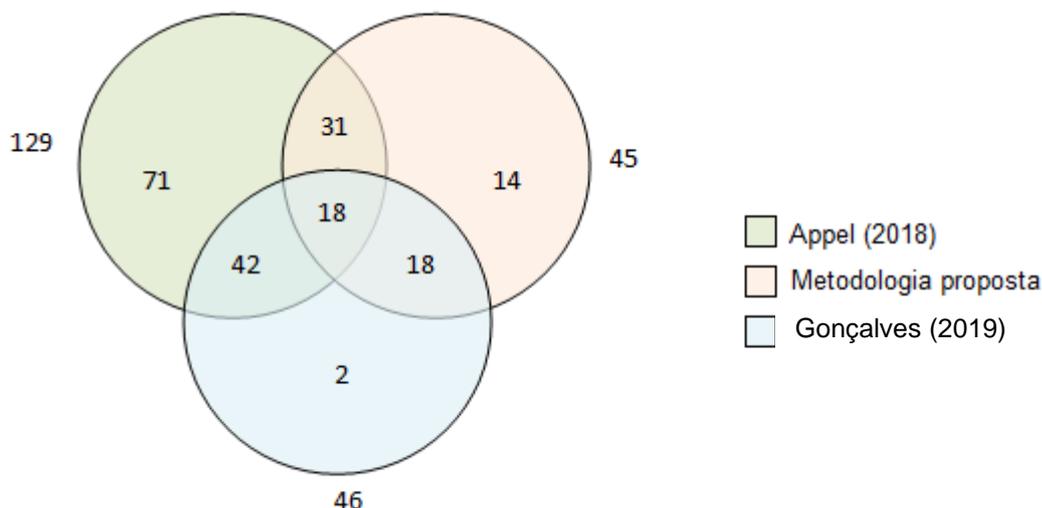


Figura 13: Comparação dos resultados das famílias de ncRNAs

A Figura 13, faz uma representação da relação entre os resultados obtidos pela comparação entre as metodologias comparadas. Pode-se observar que 18 famílias em comum foram identificadas pelas três metodologias, 31 famílias foram identificadas apenas por Appel (2018) e a metodologia proposta, 18 famílias foram identificadas apenas por Gonçalves (2019) e a metodologia proposta, e 42 famílias foram identificadas apenas por Gonçalves (2019) e Appel (2018).

A comparação entre as diferentes metodologias tem o objetivo de analisar os resultados obtidos em diferentes formas de busca, visando esclarecer os resultados obtidos por meio da análise.

6. CONCLUSÕES

Em uma análise geral, a metodologia proposta neste trabalho permitiu a identificação de um total de 355 sequências candidatas a ncRNAs distribuídas em 49 famílias. Com a remoção dos RNAs comuns (tRNA, tmRNA, 58s_RNA e 5s_RNA), a metodologia identificou 223 sequências distribuídas em 45 famílias.

A comparação da presente metodologia com as metodologias propostas por Appel (2018) e Gonçalves (2019), demonstrou uma variação significativa nas famílias identificadas. Em relação a proximidade dos resultados, a metodologia

proposta neste trabalho obteve uma proximidade maior em relação ao número de candidatos e famílias identificadas.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ABOU, E.S.; NAZAR, R.N. "Role of the 5.8S rRNA in ribosome translocation". *Nucleic Acids Research*. 25 (9): 1788–94, 1997.
- ALVARENGA, D. O.; MOREIRA, L. M.; CHANDLER, M.; VARANI, A. M. A practical guide for comparative genomics of mobile genetic elements in prokaryotic genomes. In: **Comparative Genomics**. Humana Press, New York, NY, p. 213-242, 2018.
- ALVES, S. M. A bioinformática e sua importância para a biologia molecular. **Revista Brasileira de Educação e Saúde**, v. 3, n. 4, p. 18-25, 2013.
- ANGELO, E. A.; VILAS-BOAS, G. T.; CASTRO-GÓMEZ, R. J. H. **Bacillus thuringiensis: características gerais e fermentação**. Ciências Agrárias, Londrina, v. 31, n. 4, p. 945-958, out./dez. 2010
- ARONSON, A. I. **The two faces of *Bacillus thuringiensis*: insecticidal proteins and post-exponential survival**. *Molecular microbiology*, v. 7, n. 4, p. 489-496, 1993.
- APPEL, R. J. C. "Predição, análise e validação de RNAs não codificadores (ncRNAs) em *Bacillus thuringiensis*". Dissertação de mestrado. UEL- Universidade Estadual de Londrina, 2018.
- ARRIAL, R. T. **Predição de RNAs não-codificadores no transcriptoma do fungo *Paracoccidioides brasiliensis* usando aprendizagem de máquina**. Universidade de Brasília - UnB. Dissertação de mestrado.p.15-16, 2008.
- ARRUDA, W. C. **ncRNA-Agents: Anotação de RNAs não-codificadores Baseada em Sistema Multiagente**. Universidade de Brasília – UnB. Tese, p. 4-23, 2015.
- BATISTA, R. D., PEREIRA, C. F., OLIVEIRA, A. I. T., SILVA, J. F. M. **Contaminação por *Bacillus cereus* e os riscos gerados através da intoxicação alimentar**. *Revista Desafios* – v. 5 ,n. 02, 2018 .

- BERNHART, S. H.; HOFACKER, I. L.; STADLER, P. F. **Local RNA base pairing probabilities in large sequences**. *Bioinformatics*, v. 22, n. 5, p. 614-615, 2005.
- BERRY, C.; O'NEIL, S.; BEN-DOV, E.; JONES, A. F.; MURPHY, L.; QUAIL, M. A.; GOLDE, M. T. H.; HARRIS, D.; ZARITSKY, A.; PARKHILL, J. **Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. israelensis**. *Applied and environmental microbiology*, v. 68, n. 10, p. 5082-5095, 2002.
- BONNER, E.R.; D'ELIA, J.N.; BILLIPS, B.K.; **Molecular recognition of pyr mRNA by the *Bacillus subtilis* attenuation regulatory protein PyrR**. *Nucleic Acids Research, Oxford* , v. 29, n. 23, p. 4851-4865, 2001.
- BRENNECKE, j.; ARAVIN, A. A.; STARK, A.; DUS, M.; KELLIS, M. SACHIDANANDAM, R. and HANNON, G. J. **Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila***. *Cell*, 128(6):1089-1103, 2007.
- BRIGHENTI, D. M. et al. **Eficiência do *Bacillus thuringiensis* var. *kurstaki* no controle da traça da cera *Galleria mellonella***. *Ciênc. agrotec.* vol.29 no.1 Lavras, 2005.
- BROSNAN, C. A.; VOINNET, O. **The long and the short of noncoding RNAs**. *Current opinion in cell biology*, v. 21, n. 3, p. 416-425, 2009.
- CARDOSO, T. A. O., VIEIRA, N. V. ***Bacillus anthracis* como terrorista**. *SAÚDE DEBATE* | Rio de Janeiro, v. 40, n. 107, P. 1138-1148, 2015.
- CARVER, T. HARRIS, S.R.; BERRIMAN, M.; PARKHILL, J.; MCQUILLAN, J.A. **Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data**. *Bioinformatics (Oxford, England)* p. 464, 2012
- CATANBO, M.; DEGRAVE, W. M.; DE MIRANDA, A. B. D. **Análise Comparativa de Genomas Procarióticos**. *Revista Biotecnologia Ciência & Desenvolvimento*, Ed. 37, 2007.

- CHAVES, J. Q. **Caracterização de linhagens pertencentes ao grupo *B. cereus lato sensu* isoladas no Brasil: Filogenia e avaliação do potencial toxigênico e de virulência.** Tese, Instituto Oswaldo Cruz, Rio de Janeiro. 2015.
- CORREIA, J. H. R. D. **Funcionalidades dos RNA não codificantes (ncRNA) e pequenos RNA reguladores, nos mamíferos.** Revista eletrônica de Veterinária 1695-7504, Volume VIII Número 10, Lisboa, Portugal, 2007.
- COSTA, J. C. M. **RNAs não codificadores regulatórios em bactérias do gênero *Aeromonas*.** PósGraduação em Bioinformática, Setor de Educação Profissional e Tecnológica, Universidade Federal do Paraná. Curitiba, 2017.
- COSTA, F. F. **Non-coding RNAs: new players in eukaryotic biology.** Gene, v. 357, n. 2, p. 83-94, 2005.
- DURBIN, R.; EDDY, S.R.; KROGH, A. and MITCHISON, G. **Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.** Cambridge University Press, 1998
- EDDY, S. R. Non-coding RNA genes and the modern RNA world. **Nature Reviews Genetics**, v. 2, n. 12, p. 919-929, 2001.
- FIEDORUK, K. DANILUK, T. MAHILLON, J. LESZCZYNSKA, K. and SWIECICKA, I. **Genetic Environment of cry1 Genes Indicates Their Common Origin.** *Genome Biology and Evolution* , volume 9, edição 9, páginas 2265–2275, 2017.
- FLICEK, P.; AMODE, M. R.; BARRELL, D.; BEAL, K.; BRENT, S.; CARVALHO-SILVA, D.;GIL, L. Ensembl 2012. **Nucleic acids research**, v. 40, n. D1, p. D84-D90, 2011.
- FONSECA, M. A. S. **Identificação *in silico* de ncRNAs no organismo modelo *halobacterium salinarum* NRC-1.** Tese de doutorado. Universidade de São Paulo. Ribeirão Preto, 2016.
- GARCIA, A. C. **RNAs não codificadores em *Herbaspirillum* spp.** Dissertação. Pós-graduação em Ciências-Bioquímica, setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, 29 de Setembro de 2016.

- GONÇALVES, K. C. B. **Busca e caracterização in silico de RNAs não codificadores em isolados de *Bacillus anthracis*, *Bacillus cereus* e *Bacillus thuringiensis* (*Bacillus cereus sensu lato*)**. Tese. Universidade Estadual de Londrina, Londrina, 2019.
- HAN, C. S. XIE, G.; CHALLACOMBE, J. F.; ALTHERR, M. R.; BHOTIKA, S. S.; BRUCE, D.; ... & CLELAND, C. Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. **Journal of bacteriology**, v. 188, n. 9, p. 3382-3390, 2006
- HELGASON, E., OKSTAD, O. A., CAUGANT, D. A., JOHANSEN, H. A., FOUET, A., MOCK, M., HEGNA, I., KOLSTO, A. B. ***Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* - One Species on the Basis of Genetic Evidence**. 2000, American Society for Microbiology, p. 2627–2630, 2000.
- HERTEL, J.; DE JONG, D.; MARZ, M.; ROSE, D.; TAFER, H.; TANZER, A.; SCHIERWATER, B.; STADLER, P. F. **Non-coding RNA annotation of the genome of *Trichoplax adhaerens***. Nucleic acids research, v. 37, n. 5, p. 1602-1615, 2009.
- HUART, M.; GUTTMAN, M.; FELDSER, D.; GARBER, M.; KOZIOL, M. J.; KENZELANN-BROZ, D.; KHALIL, A. M.; ZUK, O.; AMIT, I. and RABANI, M. **A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response**. Cell, 142(3):409-419, 2010
- KALANTIDIS, K.; RAO, A. L. N.; TSAGRIS, E. M.; KATSAROU, K. **Infectious long non coding RNAs**. Published by Elsevier B. V. Department of Biology, University of Crete, P.O. Box 2208, GR-71409, Heraklion, Crete, Greece, 2015.
- KERSEY, P. J.; ALLEN, J. E.; CHRISTENSEN, M.; DAVIS, P.; FALIN, L. J.; GRABMUELLER, C.; LANGRIDGE, N. Ensembl Genomes 2013: **scaling up access to genome-wide data**. Nucleic acids research, v. 42, n. D1, p. D546-D552, 2013.
- LAFONTAINE, D. L.; TOLLERVEY, D. "The function and synthesis of ribosomes". **Nature Reviews. Molecular Cell Biology**, v. 2, p. 514–520, 2011.

- LANGMEAD, B.; TRAPNELL, C.; POP, M. SALZBERG, S.L. **Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.** *Genome Biol.* 2019.
- LÉONARD, C.; CHEN, Y.; MAHILLON, J. **Diversity and differential distribution of IS231, IS232 and IS240 among *Bacillus cereus*, *Bacillus thuringiensis* and *Bacillus mycoides*.** *Microbiology*, v. 143, n. 8, p. 2537-2547, 1997.
- LOEZA-LARA, P. D.; BENINTENDE, G.; COZZI, J.; OCHOA-ZARZOSA, A.; BAIZABAL-AGUIRRE, V. M.; VALDEZ-ALARCÓN, J. J.; LÓPEZ-MEZA, J. E. **The plasmid pBMBt1 from *Bacillus thuringiensis* subsp. darmstadiensis (INTA Mo14-4) replicates by the rolling-circle mechanism and encodes a novel insecticidal crystal protein-like gene.** *Plasmid*, v. 54, n. 3, p. 229-240, 2005.
- LIU, Y. LAI, Q. GOKER, M. P, J. WANG, M. SUN, Y. WANG, L. SHAO, Z. **Genomic insights into the taxonomic status of the *Bacillus cereus* group.** *Nature, Scientific Reports* volume 5, Article number: 14082. 2015.
- LIU, L.; LI, Y.; LI, S.; HU, N.; HE, Y.; PONG, R.; LIN, D; LU, L.; LAW, M. Comparison of next-generation sequencing systems. **BioMed Research International**, v. 2012.
- MAHILLON, J.; REZSÖHAZY, R.; HALLET, B.; DELCOUR, J. **IS231 and other *Bacillus thuringiensis* transposable elements: A review.** *Genetica*, v. 93, n. 1-3, p. 13-26, 1994.
- MENDELL, J. T. **Micrnas: critical regulators of development, cellular physiology and malignancy.** *Cell Cycle*, 4(9):1179-1184, 2005.
- MENDES, R. E. **Caracterização dos elementos genéticos móveis responsáveis pela disseminação de genes associados a resistência bacteriana em em *Pseudomonas* spp. Isoladas na América Latina.** Tese. Universidade Federal de São Paulo – Escola Paulista de Medicina, 2005.

- MARCELINO, A. F. **Bacillus thuringiensis como bioinseticida: uma avaliação o estado da arte e bioensaio**. Dissertação. Universidade Tecnológica Federal do Paraná, Dois Vizinhos, 2019.
- MISHRA, R. **Biosafety of *Bt*-crop in soil ecosystem: A review**. IJAR, v. 3, n. 7, p. 107-119, 2017. MATSUDA, H.; YOSHIKAWA, T.; TABE, T.; KISHINAMI, R. & HASHIMOTO, A. **On the Implementation of a Phylogenetic Tree Database**. IEEE, 42-45, 1999.
- NAWROCKI. E.P.; BURGE. S. W.; BATEMAN. A.; et al **.Rfam 12.0: updates to the RNA families database**. Nucleic Acids Research, London, v. 43, n/d, p. 130-137, nov. 2014.
- NADAL, A. L. **Busca e caracterização *in silico* de RNAs não codificadores em isolados de *Bacillus anthracis*, *Bacillus cereus* e *Bacillus thuringiensis* (*Bacillus cereus sensu lato*)**. Dissertação de mestrado. UEL. Londrina, PR, p. 1-92, 2015.
- OKINAKA, R. T., KEIM, P. **The Phylogeny of *Bacillus cereus sensu lato***. Edited by P. Eichenberger and A. Driks © 2016 American Society for Microbiology, Washington, DC, 2016.
- PALMA, L.; MUÑOZ, D.; BERRY, C.; MURILLO, J.; CABALLERO, P. ***Bacillus thuringiensis* toxins: an overview of their biocidal activity**. Toxins, v. 6, n. 12, p. 3296-3325, 2014.
- PAVON-ETERNOD, M.; GOMES, S.; GESLAIN, R.; DAI, Q.; ROSNER, M. R.; and PAN, T. **tRNA over-expression in breast cancer and functional consequences**. Nucleic Acids Research, 37(21):7268-7280, 2009.
- PASCHOAL. A. R.; MARACAJA-COUTINHO. V.; Setubal. J.C.; SIMÕES. Z. L. P.; ALMEIDA. S. V.; DURHAM. A. M. **Non-coding transcription characterization and annotation: A guide and web resource for non-coding RNA Databases**. RNA Biology, Georgetown, v. 9, n. 3, p. 274–282, 2012.
- PONTING, C. P.; OLIVER, P. L. and REIK, W. **Evolution and functions of long noncoding RNAs**. Cell, 136(4):629-641, 2009.

- PRADO, O. G., **Computação evolutiva empregada na reconstrução de árvores filogenéticas**. Dissertação. Universidade Estadual de Campinas, Campinas, SP. p. 1-174, 2001.
- PROSDÓCIMI, F. MOREIRA, L. M.; **Ciências genômicas: fundamentos e aplicações**. Genômica comparativa. Livro, v. 1, p. 81-99, 2015.
- PROSDÓCIMI, F.; CERQUEIRA, G. C.; BINNECK, E.; SILVA, A.F. FILHO, F.C.; PETRETSKI, J. H.; KESSEDJIAN, J. L.; CAMARGO, L. P. 2002. **Bioinformática: manual do usuário**. Artigo. Biotecnologia Ciência & Desenvolvimento - nº 29, 2002.
- RABINOVITCH, L.; LOURENÇO, M. C. **Bacillus anthracis, pós e bioterrorismo**. Rio de Janeiro, v. 39, n. 1, p. 49, 2003.
- RASKO, D. A.; ALTHERR, M. R.; , HAN C. S.; , RAVEL, J. **Genomics of the Bacillus cereus group of organisms**. FEMS Microbiology Reviews. First published: 09 January 2006. <https://doi.org/10.1016/j.fmrre.2004.12.005>
- REPOILA, F. DARFEUILLE, F. **Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects**. Biol. Volume 101 (2) | Pages 117–131, 2009.
- ROH, J. Y.; CHOI, J. Y.; LI, M. S.; JIN, B. R.; JE, Y. H. *Bacillus thuringiensis* as a specific, safe, and effective tool for insect pest control. **Journal of microbiology and biotechnology**, v. 17, n. 4, p. 547, 2007.
- ROSSI, G. A. M., AGUILAR, C. E. G., SILVA, H. O., VIDAL, A. M. C. *Bacillus cereus* group: genetic aspects related to food safety and dairy processing. **Arquivos do Instituto Biológico. Arq. Inst. Biol. vol.85 São Paulo Epub Nov 01, 2018**.
- SAADAQUI, I.; AL-THANI, R.; AL-SAAD, F.; HASSAN, N. B. B.; ABDELKEFI-MESRATI, L.; SCHULTZ, P.; ROUIS, S.; JAOUA, S. **Characterization of tunisian Bacillus thuringiensis strains with abundance of kurstaki**

subspecies harbouring insecticidal activities against the lepidopteran insect *Ephestia kuehniella*. *Current microbiology*, v. 61, n. 6, p. 541-548, 2010.

SEPÚLVEDA, G. **Enciclopédia temática. Genômica comparative.** Set. 2015. Acesso em: 16 de setembro de 2019.

SCHNEPF, E.; CRICKMORE, N.; RIE, J. V.; LERECLUS, D.; BAUM, J.; FEITELSON, J.; ZEIGLER, D. R.; DEAN, D. H. *Bacillus thuringiensis* and its pesticidal crystal proteins. **Microbiology and Molecular Biology Review**, Washington, s.n., v. 62, p. 775-806, 1998.

SHANMUGAM, A.; NAGARAJAN, A.; PRAMANAYAGAM, S. Non-coding DNA—a brief review. **Journal of Applied Biology & Biotechnology** Vol, v. 5, n. 05, p. 42-47, 2017.

SHEPPARD, A. E.; POEHLEIN, A.; ROSENSTIEL, P.; LIESEHANG, H.; SCHULENBURG, H. **Complete Genome Sequence of Bacillus thuringiensis Strain 407 Cry.** Artigo, v. 1, 2013.

SMILLIE, C.; GARCILLÁN-BARCIA, M. P.; FRANCIÁ, M. V.; ROCHA, E. P.; DE LA CRUZ, F. Mobility of plasmids. **Microbiology and Molecular Biology Reviews**, v. 74, n. 3, p. 434-452, 2010.

SONG, D.; YANG, Y.; YU, B.; ZHENG, B.; DENG, Z.; LU, B. L.; CHEN, X.; JIANG, T. Computational prediction of novel non-coding RNAs in *Arabidopsis thaliana*. **BMC bioinformatics**, v. 10, n. 1, p. S36, 2009.

SOUZA, M. V. N.; ALMEIDA, M. V.; SILVA, A. D.; COURI, M. R. C. **Ciprofloxacina, uma importante fluorquinolona no combate ao antraz.** *Rev. Bras. Farm.*, 85(1), 2004.

STEVENS, M. A.; TASARA, T.; KLUMPP, J.; STEPHAN, R.; SCHULZ, M. E.; JOHLER, S. **Whole-genome-based phylogeny of Bacillus cytotoxicus reveals different clades within the species and provides clues on ecology and evolution.** Artigo. *Nature*, 2019.

- SUZUKI, M.; HAYASHIZAKI, Y. **Mouse-centric comparative transcriptomics of protein coding and non-coding RNAs**. *Bioessays*, v. 26, n. 8, p. 833-843, 2004.
- TAMEZ-GUERRA, P.; GALÁN-WONG, L. J.; MEDRADO-ROLDÁN, H.; GARCÍA-GUTIÉRREZ, C.; RODRÍGUEZ-PADILLA, C.; GÓMEZ-FLORES, R. A.; TAMEZ-GUERRA, R. S. **Bioinsecticidas: su empleo, producción y comercialización em México, Ciência UANL, San Nicolás de los Garzas**, v. 4, n. 2. p. 143-152, 2001.
- TRAVASSOS, G. H. BARROS, M. O. 2003. Contributions of In Virtuo and In Silico Experiments for the Future of Empirical Studies in Software Engineering. In **Proceedings of 2nd Workshop Serie Empirical Software Engineering**. Rome: RESEARCH GATE, p. 177-130, 2004.
- VIEIRA, L. M. **Classificação de RNAs não-codificadores Longos Intergênicos usando Máquina de Vetores de Suporte: um Estudo de Caso para a Cana-de-açúcar**. Trabalho de conclusão. Universidade de Brasília, Brasília, p.4-29, 2016.
- VILAS-BÔAS, G. T.; PERUCA, A. P. S.; ARANTES, O. M. N. **Biology and taxonomy of *Bacillus cereus*, *Bacillus anthracis* and *Bacillus thuringiensis***. *Canadian Journal Microbiology*, Ottawa, v. 53, n. 6, p. 673-687, jun. 2007.
- WATERS, L. S.; STORZ, G. **Regulatory RNAs in bacteria**. *Cell*, v. 136, n. 4, p. 615-628, 2009.
- WENFEI, Z.; HONGPING, W.; ZHIXIA, X.; YINGHONG, J.; ZHONGQI, W.; JIANGZHAO, Q.; ... NING, W. **Mining new *Bt* insecticidal protein genes resource using Illumina sequencing technology**. *Bt Research*, v. 5, 2014.
- XUE, C.; LI, F. **Finding noncoding RNA transcripts from low abundance expressed sequence tags**. *Cell research*, v. 18, n. 6, p. 695-700, 2008.
- ZENGUEL, J.M.; LINDAHL, L. Diverse mechanisms for regulating ribossomal protein synthesis in *Escherichia coli*. **Progress in Nucleic Acid Research and Molecular Biology**, New York, v.47, n--, p. 331-370, set. 1994.

ZHONG, C.; PENG, D.; YE, W.; CHAI, L.; QI, J.; YU, Z.; RUAN, L.; SUN, M.
Determination of plasmid copy number reveals the total plasmid DNA amount is greater than the chromosomal DNA amount in *Bacillus thuringiensis* YBT-1520. PloS One, v. 6, n. 1, p. e16025, 2011.

APÊNDICES

1. Protocolo utilizado na identificação de ncRNAs

Especificações de Software

- o computador utilizado para as análises utilizou o Sistema Operacional Windows 10 *Home Single Language* e o Sistema Operacional Ubuntu 16.04.4.

Especificações de Hardware

- notebook DELL
- Processador Intel® Core™ i3-5005 CPU @ 2.00GHz 2.00GHz
- Memória RAM 4,00 GB
- Sistema Operacional de 64 bits, processador com base em 64 bits

2. Extração das sequências de RNA utilizando Artemis

Abrir o Artemis e selecionar no menu a opção “File” – “Open”:

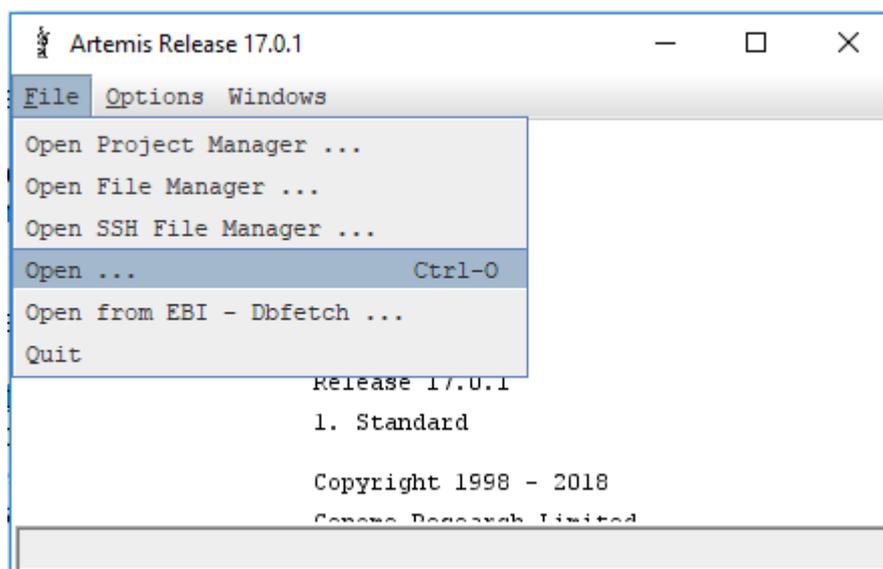


Figura 14: Abrir arquivo GenBank (.gbff) no Artemis

Selecionar arquivo GenBank (.gbff):

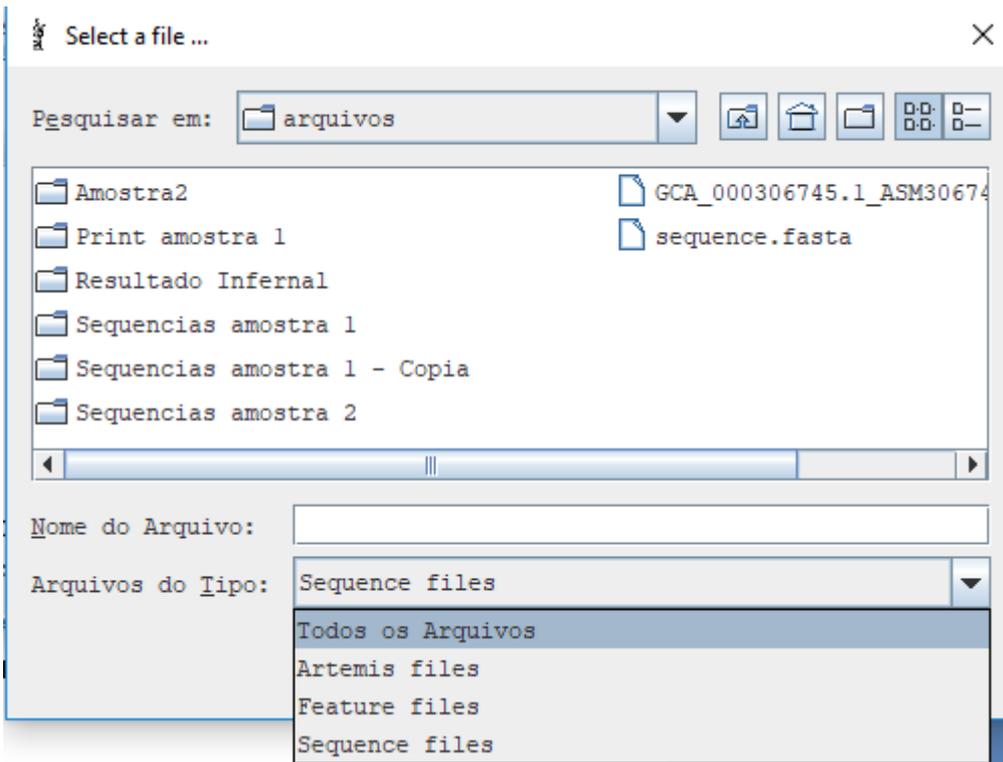


Figura 15: Selecionar arquivo GenBank (.gbff) no Artemis

Com o arquivo GenBank já aberto selecionar no menu “File” – “Read BAM” e selecionar arquivo com a extensão “.sorted.bam”:



Figura 16: Abrir arquivo de alinhamento .sorted.bam



Figura 17: Janela de visualização do genoma no Artemis

Para obter a sequência FASTA da região desejada clicar com o botão direito sobre a região e selecionar “View” – “Bases” – “Bases Of Selection As FASTA”:

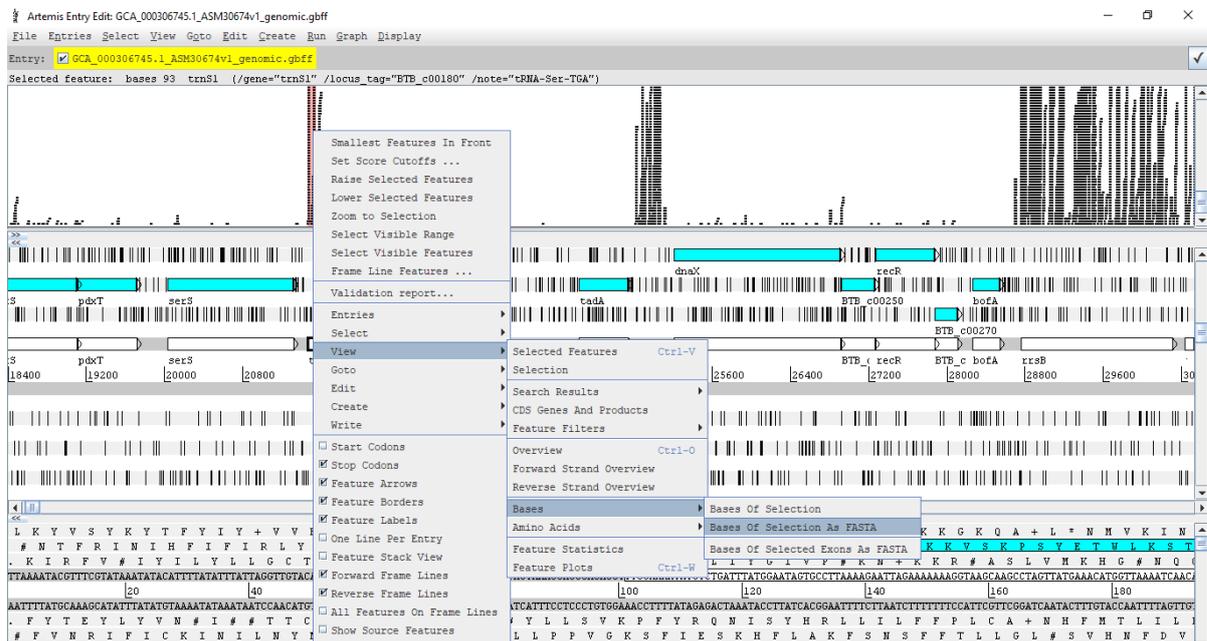


Figura 18: Extrair sequência FASTA no Artemis

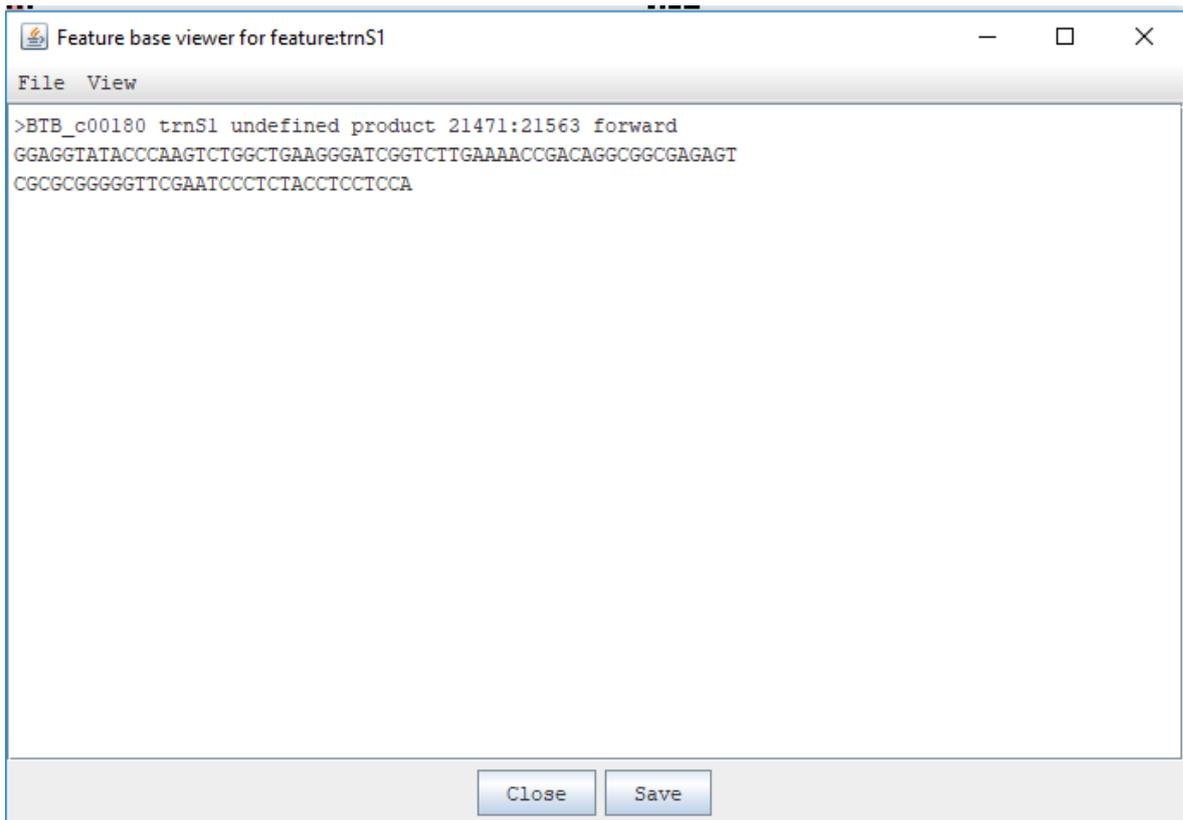


Figura 19: Exemplo de sequência FASTA extraída com Artemis

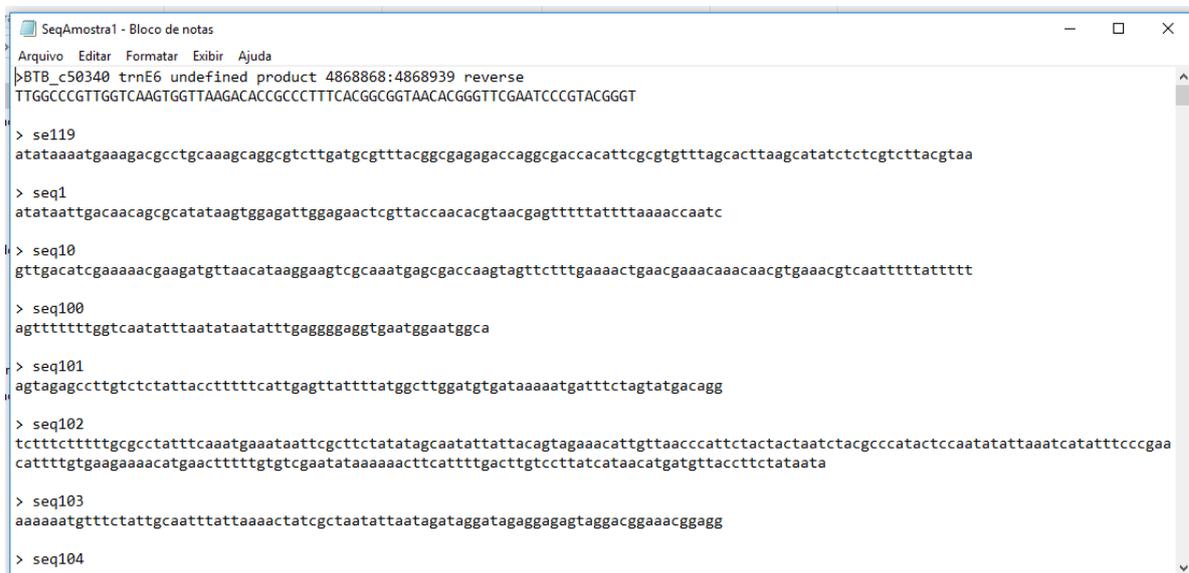


Figura 20: Arquivo multiFASTA com todas as sequências da Amostra 1

3. Software Infernal e Banco de Dados Rfam

Comando:

```
cmsearch -tblout Result_Infernal.tbl /home/Viviane/arquivos/Bacilalles.cm
SeqAmostr1 >stdout.txt 2>stderr.txt &
```

Tabela 6: Possíveis seqüências candidatas para Amostra 1

Nome	Familia	Posicao	Score	E-value	Coordordenadas
BTB_c50560	5S_rRNA	+	69.3	4.2e-16	4870836:4870950
BTB_c03360	5S_rRNA	+	69.3	4.2e-16	310188:310302
BTB_c03080	5S_rRNA	+	68.6	6.3e-16	271700:271814
BTB_c05370	5S_rRNA	+	65.3	4.6e-15	530111:530225
BTB_c07930	5S_rRNA	+	65.3	4.6e-15	788873:788987
BTB_c02850	5_8S_rRNA	+	45.7	3.3e-12	245612:248531
BTB_c50390	tRNA	+	77.2	1.8e-19	4869297:4869372
BTB_c50370	tRNA	+	72.3	4.3e-18	4869126:4869202
BTB_c50400	tRNA	+	71.4	7.6e-18	4869383:4869458
BTB_c50410	tRNA	+	71.3	8.5e-18	4869473:4869548
BTB_c02720	tRNA	+	71.1	9.5e-18	242657:242731
BTB_c02740	tRNA	+	69.0	3.6e-17	242834:242908
BTB_c07970	tRNA	+	67.4	1,00E-16	789250:789325
BTB_c02820	tRNA	+	67.0	1.3e-16	243687:243763
BTB_c46380	tRNA	+	66.3	2.1e-16	4493107:4493183
BTB_c02750	tRNA	+	65.7	3.1e-16	242913:242988
BTB_c50430	tRNA	+	64.5	6.6e-16	4869637:4869713
BTB_c50420	tRNA	+	63.4	1.4e-15	4869558:4869633
BTB_c56660	tRNA	+	63.4	1.4e-15	5479700:5479775
BTB_c02880	tRNA	+	63.4	1.4e-15	248790:248865
BTB_c08110	tRNA	+	62.2	2.8e-15	790568:790644
BTB_c50380	tRNA	+	62.1	3.1e-15	4869213:4869283
BTB_c50530	tRNA	+	61.7	3.9e-15	4870591:4870666
BTB_c08080	tRNA	+	61.2	5.7e-15	790322:790397
BTB_c08120	tRNA	+	61.1	5.8e-15	790653:790727
seq3	tRNA	+	60.7	7.7e-15	21471:21563
BTB_c44190	tRNA	+	60.4	9.1e-15	4283075:4283148
BTB_c50340	tRNA	+	60.1	1.1e-14	4868868:4868939
BTB_c56670	tRNA	+	60.1	1.1e-14	5479797:5479868
BTB_c05580	tRNA	+	59.4	1.8e-14	535974:536047
BTB_c05500	tRNA	+	59.1	2.2e-14	531348:531422
BTB_c07990	tRNA	+	59.1	2.2e-14	789451:789525
BTB_c50460	tRNA	+	58.4	3.4e-14	4869933:4870009
BTB_c50480	tRNA	+	58.0	4.3e-14	4870118:4870191
BTB_c08020	tRNA	+	57.0	8.3e-14	789718:789791
BTB_c05460	tRNA	+	55.8	1.8e-13	530945:531028
BTB_c07950	tRNA	+	55.8	1.8e-13	789076:789159
BTB_c07940	tRNA	+	55.6	2,00E-13	788992:789067
BTB_c05470	tRNA	+	54.7	3.6e-13	531036:531109
BTB_c08100	tRNA	+	54.7	3.6e-13	790487:790557
BTB_c50470	tRNA	+	54.2	5,00E-13	4870030:4870102
BTB_c50490	tRNA	+	53.9	6,00E-13	4870202:4870275
BTB_c05430	tRNA	+	53.3	9.1e-13	530680:530755
BTB_c08070	tRNA	+	53.3	9.1e-13	790238:790313

BTB_c05480	tRNA	+	51.7	2.5e-12	531129:531204
BTB_c02800	tRNA	+	50.6	5.2e-12	243427:243512
BTB_c05390	tRNA	+	49.2	1.3e-11	530310:530401
BTB_c07980	tRNA	+	47.6	3.5e-11	789341:789421
BTB_c08000	tRNA	+	46.8	5.9e-11	789542:789630
BTB_c13160	tRNA	+	46.1	9.3e-11	1278973:1279045
BTB_c08060	tRNA	+	41.7	1.5e-09	790167:790234
seq77	RNaseP_bact_a	+	56.7	2.4e-18	1533976:1534389
seq77	RNaseP_bact_b	+	317.7	1.5e-95	1533976:1534389
seq78	6S	+	109.4	8.3e-26	1550683:1551186
seq78	6S	+	93.2	7.4e-22	1550683:1551186
seq1151	tmRNA	-	178.8	3,00E-54	5109055:5109435
BTB_c05460	tmRNA	+	19.2	1.9e-05	530945:531028
BTB_c07950	tmRNA	+	19.2	1.9e-05	789076:789159
BTB_c50340	tmRNA	+	11.4	0.0045	4868868:4868939
BTB_c02740	tmRNA	+	11.4	0.0045	242834:242908
BTB_c56670	tmRNA	+	11.4	0.0045	5479797:5479868
seq73	FMN	+	121.7	1.2e-31	1467400:1467690
seq13	SAM	+	21.7	0.0011	182853:183015
seq2	Bacteria_small_SRP	+	80.0	8.1e-21	24820:25152
seq4	Bacteria_small_SRP	+	80.0	8.1e-21	24827:25144
BTB_c50580	SSU_rRNA_bacteria	+	1633.1	0	4874127:4875666
BTB_c03060	SSU_rRNA_bacteria	+	1632.9	0	267037:268576
seq70	T-box	+	88.6	5.6e-22	1416868:1417239
seq117	T-box	+	58.2	4.1e-14	1099511:1099754
seq128	T-box	+	41.5	8.2e-10	4730764:4730973
seq67	T-box	+	31.9	2.5e-07	1247347:1247406
seq81	T-box	+	15.0	0.0058	1759657:1759821
seq85	Glycine	-	62.1	1.7e-14	2307556:2307900
seq128	PyrR	+	25.3	0.0044	4730764:4730865
seq9	L10_leader	+	85.5	2.1e-15	111607:111813
seq13	L10_leader	+	21.6	0.005	182869:182979
seq123	L20_leader	+	43.5	4.7e-09	925637:925799
seq28	c-di-GMP-I	+	72.1	1.1e-15	423400:423582
se119	BsrF	-	99.2	2.3e-21	4420276:4420395
seq3	tRNA-Sec	+	43.0	3.3e-09	21433:21651
BTB_c50390	tRNA-Sec	+	34.0	7.2e-07	4869297:4869372
BTB_c50400	tRNA-Sec	+	31.1	4,00E-06	4869383:4869458
BTB_c50410	tRNA-Sec	+	30.5	5.5e-06	4869473:4869548
BTB_c50530	tRNA-Sec	+	30.2	6.8e-06	4870591:4870666
BTB_c05470	tRNA-Sec	+	28.7	1.7e-05	531036:531109
BTB_c08100	tRNA-Sec	+	28.7	1.7e-05	790487:790557
BTB_c02880	tRNA-Sec	+	27.2	4,00E-05	248790:248865
BTB_c50420	tRNA-Sec	+	27.2	4,00E-05	4869558:4869633
BTB_c56660	tRNA-Sec	+	27.2	4,00E-	5479700:5479775

05					
BTB_c02740	tRNA-Sec	+	27.2	4.1e-05	242834:242908
BTB_c50370	tRNA-Sec	+	26.7	5.5e-05	4869126:4869202
BTB_c02750	tRNA-Sec	+	26.3	6.8e-05	242913:242988
BTB_c08000	tRNA-Sec	+	26.3	7.2e-05	789542:789630
BTB_c07970	tRNA-Sec	+	26.2	7.4e-05	789250:789325
BTB_c50430	tRNA-Sec	+	25.9	8.8e-05	4869637:4869713
BTB_c07980	tRNA-Sec	+	24.4	0.00022	789341:789421
BTB_c02720	tRNA-Sec	+	23.8	0.00031	242657:242731
BTB_c50340	tRNA-Sec	+	23.4	0.00039	4868868:4868939
BTB_c56670	tRNA-Sec	+	23.4	0.00039	5479797:5479868
BTB_c50480	tRNA-Sec	+	23.2	0.00045	4870118:4870191
BTB_c02820	tRNA-Sec	+	23.2	0.00045	243687:243763
BTB_c44190	tRNA-Sec	+	23.1	0.00047	4283075:4283148
BTB_c08020	tRNA-Sec	+	21.7	0.0011	789718:789791
BTB_c05460	tRNA-Sec	+	21.5	0.0012	530945:531028
BTB_c07950	tRNA-Sec	+	21.5	0.0012	789076:789159
BTB_c05500	tRNA-Sec	+	21.5	0.0012	531348:531422
BTB_c07990	tRNA-Sec	+	21.5	0.0012	789451:789525
BTB_c08120	tRNA-Sec	+	21.4	0.0013	790653:790727
BTB_c50490	tRNA-Sec	+	20.8	0.0018	4870202:4870275
BTB_c08080	tRNA-Sec	+	20.7	0.002	790322:790397
BTB_c08110	tRNA-Sec	+	20.1	0.0029	790568:790644
BTB_c05480	tRNA-Sec	+	19.5	0.0042	531129:531204
BTB_c05390	tRNA-Sec	+	19.0	0.0055	530310:530401
BTB_c02800	tRNA-Sec	+	18.9	0.0058	243427:243512
BTB_c50460	tRNA-Sec	+	18.5	0.0075	4869933:4870009
seq2	Bacteria_large_SRP	+	242.0	5.1e-47	24799:25143
seq4	Bacteria_large_SRP	+	242.0	5.1e-47	24936:24967
BTB_c50580	SSU_rRNA_eukarya	+	743.6	1.3e-225	4874127:4875666
BTB_c03060	SSU_rRNA_eukarya	+	741.2	6.5e-225	267037:268576
seq76	ncr1175	-	23.5	0.00087	1530293:1530352
seq1152	ncr1241	+	20.4	0.0084	5209527:5209635
BTB_c02850	LSU_rRNA_archaea	+	1996.4	0	245612:248531
BTB_c02850	LSU_rRNA_bacteria	+	3003.9	0	245612:248531
BTB_c02850	LSU_rRNA_eukarya	+	1294.9	0	245612:248531
seq76	rliE	-	21.6	0.004	1530293:1530352
seq76	rliE	+	21.1	0.0051	1530293:1530352
seq67	rli33	+	11.7	0.0093	1247353:1247472
seq13	RsaI	-	15.0	0.005	5303639:5303729
seq162	sau-27	+	22.0	0.0075	5449554:5449662
seq76	sau-5971	+	22.7	0.0015	1530293:1530352
seq75	sau-5971	+	20.2	0.0054	1496285:1496315

Tabela 7: Possíveis sequências candidatas para Amostra 2

Nome	Familia	Posicao	Score	E-value	Coordenadas
BTB_c00110	5S_rRNA	+	69.3	5.6e-16	13906:14020
BTB_c02860	5S_rRNA	+	69.3	5.6e-16	248584:248698
BTB_c03130	5S_rRNA	+	69.3	5.6e-16	278935:279049
BTB_c03360	5S_rRNA	+	69.3	5.6e-16	310188:310302
BTB_c00330	5S_rRNA	+	69.3	5.6e-16	33650:33764
BTB_c00880	5S_rRNA	+	68.7	8.1e-16	86758:86872
BTB_c01760	5S_rRNA	+	68.7	8.1e-16	155638:155752
BTB_c03080	5S_rRNA	+	68.6	8.4e-16	271700:271814
BTB_c05370	5S_rRNA	+	65.3	6.1e-15	530111:530225
BTB_c07930	5S_rRNA	+	65.3	6.1e-15	788873:788987
BTB_c00990	5_8S_rRNA	+	45.7	4.5e-12	89445:92364
BTB_c03070	5_8S_rRNA	+	45.7	4.5e-12	268730:271649
BTB_c03120	5_8S_rRNA	+	45.7	4.5e-12	275963:278882
BTB_c03350	5_8S_rRNA	+	45.7	4.5e-12	307217:310136
BTB_c07920	5_8S_rRNA	+	45.7	4.5e-12	785850:788769
BTB_c50390	tRNA	+	77.2	2.5e-19	4869297:4869372
BTB_c01830	tRNA	+	77.2	2.5e-19	156335:1564
BTB_c02790	tRNA	+	77.2	2.5e-19	243337:243409
BTB_c00080	tRNA	+	72.3	5.7e-18	10692:10768
BTB_c00900	tRNA	+	72.1	6.5e-18	86956:87031
BTB_c50540	tRNA	+	72.1	6.5e-18	4870677:4870752
BTB_c01820	tRNA	+	69.7	3.2e-17	156255:156329
BTB_c00930	tRNA	+	69.1	4.4e-17	87245:87319
BTB_c01840	tRNA	+	69.1	4.4e-17	156416:156487
BTB_c01790	tRNA	+	69.0	4.8e-17	155934:156008
BTB_c00690	tRNA	+	69.0	4.8e-17	64102:64173
BTB_c00960	tRNA	+	67.0	1.8e-16	87510:87583
BTB_c46380	tRNA	+	66.3	2.8e-16	4493107:4493183
BTB_c01810	tRNA	+	65.9	3.7e-16	156106:156189
BTB_c01800	tRNA	+	65.7	4.1e-16	156014:156089
BTB_c02750	tRNA	+	65.7	4.1e-16	242913:242988
BTB_c02870	tRNA	+	64.5	8.8e-16	248710:248786
BTB_c00090	tRNA	+	63.1	2.2e-15	10777:10852
BTB_c05510	tRNA	+	63.0	2.4e-15	531437:531507
BTB_c08110	tRNA	+	62.2	3.8e-15	790568:790644
BTB_c02810	tRNA	+	62.0	4.5e-15	243606:243682
BTB_c05400	tRNA	+	61.5	6.3e-15	530419:530493
BTB_c08130	tRNA	+	61.5	6.3e-15	790729:790800
BTB_c08090	tRNA	+	61.4	6.6e-15	790408:790483
BTB_c08080	tRNA	+	61.2	7.6e-15	790322:790397
BTB_c08120	tRNA	+	61.1	7.8e-15	790653:790727
BTB_c00180	tRNA	+	60.7	1,00E-14	21471:21563
BTB_c02730	tRNA	+	60.4	1.3e-14	242735:242825

BTB_c05380	tRNA	+	60.2	1.4e-14	530233:530307
BTB_c56670	tRNA	+	60.1	1.5e-14	5479797:5479868
BTB_c05490	tRNA	+	59.7	2,00E-14	531268:531342
BTB_c07960	tRNA	+	59.7	2,00E-14	789171:789245
BTB_c05580	tRNA	+	59.4	2.4e-14	535974:536047
BTB_c05500	tRNA	+	59.1	2.9e-14	531348:531422
BTB_c00680	tRNA	+	58.4	4.6e-14	64015:64088
seq250	tRNA	-	57.2	9.9e-14	5128504:5128684
BTB_c08060	tRNA	+	57.1	1,00E-13	790158:790234
BTB_c08020	tRNA	+	57.0	1.1e-13	789718:789791
BTB_c05460	tRNA	+	55.8	2.5e-13	530945:531028
BTB_c05410	tRNA	+	55.6	2.7e-13	530498:530573
BTB_c05470	tRNA	+	54.7	4.8e-13	531036:531109
BTB_c08100	tRNA	+	54.7	4.8e-13	790487:790557
BTB_c08030	tRNA	+	53.0	1.5e-12	789807:789879
BTB_c08010	tRNA	+	51.9	2.9e-12	789634:789707
BTB_c05480	tRNA	+	51.7	3.4e-12	531129:531204
BTB_c08040	tRNA	+	50.6	6.9e-12	789900:789992
BTB_c08050	tRNA	+	50.6	6.9e-12	790046:790138
BTB_c02800	tRNA	+	50.6	7,00E-12	243427:243512
BTB_c05390	tRNA	+	49.2	1.7e-11	530310:530401
BTB_c50520	tRNA	+	48.7	2.3e-11	4870488:4870568
BTB_c13160	tRNA	+	46.1	1.2e-10	1278973:1279045
seq202	RNaseP_bact_a	+	56.7	3.3e-18	1533959:1534447
seq202	RNaseP_bact_b	+	317.7	2,00E-95	1533959:1534447
seq203	6S	+	109.4	1.1e-25	1550713:1551192
seq203	6S	+	93.2	9.9e-22	1550713:1551192
seq249	tmRNA	-	178.8	4.1e-54	5109016:5109591
BTB_c01810	tmRNA	+	19.4	2.3e-05	156106:156189
BTB_c05460	tmRNA	+	19.2	2.5e-05	530945:531028
BTB_c00090	tmRNA	+	17.7	7.3e-05	10777:10852
BTB_c05400	tmRNA	+	11.5	0.006	530419:530493
BTB_c56670	tmRNA	+	11.4	0.0061	5479797:5479868
BTB_c01790	tmRNA	+	11.4	0.0061	155934:156008
BTB_c08130	tmRNA	+	11.4	0.0061	790729:790800
BTB_c00690	tmRNA	+	11.4	0.0061	64102:64173
seq232	FMN	+	118.8	1,00E-30	4130011:4130193
seq105	FMN	+	83.2	4.7e-21	1467445:1467660
seq57	TPP	+	55.6	1.2e-11	429136:429345
seq252	TPP	-	52.5	6.6e-11	5212591:5212779
seq242	SAM	-	35.3	3.9e-07	4755262:4755354
seq6	Bacteria_small_SRP	+	80.0	1.1e-20	24776:25165
BTB_c03160	SSU_rRNA_bacteria	+	1634.1	0	282956:284495
BTB_c03340	SSU_rRNA_bacteria	+	1634.1	0	305524:307063
BTB_c00070	SSU_rRNA_bacteria	+	1633.8	0	9017:10556
BTB_c07910	SSU_rRNA_bacteria	+	1633.1	0	784157:785696

BTB_c00980	SSU_rRNA_bacteria	+	1632.9	0	87752:89291
BTB_c02840	SSU_rRNA_bacteria	+	1630.9	0	243920:245459
seq45	SSU_rRNA_bacteria	+	21.7	4.2e-06	526908:527081
seq104	T-box	+	88.6	7.6e-22	1416874:1417191
seq99	T-box	-	48.7	1.6e-11	1197178:1197237
seq215	T-box	-	24.8	2.4e-05	2999269:2999376
seq101	T-box	+	24.4	3,00E-05	1247362:1247421
seq207	T-box	+	17.0	0.0025	1759657:1759803
seq208	Glycine	-	62.1	2.3e-14	2307578:2307868
seq203	PyrR	+	28.9	0.0014	1550698:1551210
seq232	PreQ1	-	17.4	0.0098	4129975:4130184
seq17	L10_leader	+	85.5	2.8e-15	111645:111815
seq223	L10_leader	-	22.3	0.0048	1574130:1574282
seq56	c-di-GMP-I	+	72.1	1.5e-15	423370:423585
seq234	BsrF	-	123.4	7.8e-27	4420273:4420419
seq100	RsaE	+	69.6	4.2e-13	1215730:1215873
BTB_c00180	tRNA-Sec	+	43.0	4.4e-09	21471:21563
BTB_c08050	tRNA-Sec	+	34.9	5.4e-07	790046:790138
BTB_c08040	tRNA-Sec	+	34.4	7.2e-07	789900:789992
BTB_c50390	tRNA-Sec	+	34.0	9.6e-07	4869297:4869372
BTB_c02730	tRNA-Sec	+	30.9	5.8e-06	242735:242825
BTB_c01810	tRNA-Sec	+	30.3	8.6e-06	156106:156189
BTB_c02790	tRNA-Sec	+	30.1	9.9e-06	243337:243409
BTB_c01830	tRNA-Sec	+	30.1	9.9e-06	156335:156410
BTB_c50540	tRNA-Sec	+	29.6	1.3e-05	4870677:4870752
BTB_c00900	tRNA-Sec	+	29.6	1.3e-05	86956:87031
BTB_c05470	tRNA-Sec	+	28.7	2.3e-05	531036:531109
BTB_c08100	tRNA-Sec	+	28.7	2.3e-05	790487:790557
BTB_c00690	tRNA-Sec	+	27.2	5.5e-05	64102:64173
BTB_c01790	tRNA-Sec	+	27.2	5.5e-05	155934:156008
BTB_c00080	tRNA-Sec	+	26.7	7.3e-05	10692:10768
BTB_c01800	tRNA-Sec	+	26.3	9.1e-05	156014:156089
BTB_c02750	tRNA-Sec	+	26.3	9.1e-05	242913:242988
BTB_c02870	tRNA-Sec	+	25.9	0.00012	248710:248786
BTB_c02810	tRNA-Sec	+	25.3	0.00017	243606:243682
BTB_c50520	tRNA-Sec	+	24.4	0.00029	4870488:4870568
BTB_c00930	tRNA-Sec	+	24.0	0.00038	87245:87319
BTB_c01840	tRNA-Sec	+	24.0	0.00038	156416:156487
BTB_c05510	tRNA-Sec	+	23.6	0.00046	531437:531507
BTB_c56670	tRNA-Sec	+	23.4	0.00053	5479797:5479868
BTB_c00960	tRNA-Sec	+	23.2	0.0006	87510:87583
BTB_c01820	tRNA-Sec	+	23.2	0.0006	156255:156329
BTB_c05380	tRNA-Sec	+	22.8	0.00077	530233:530307
BTB_c08090	tRNA-Sec	+	22.0	0.0013	790408:790483
BTB_c08020	tRNA-Sec	+	21.7	0.0014	789718:789791
BTB_c05460	tRNA-Sec	+	21.5	0.0016	530945:531028

BTB_c05500	tRNA-Sec	+	21.5	0.0017	531348:531422
BTB_c08120	tRNA-Sec	+	21.4	0.0018	790653:790727
BTB_c08060	tRNA-Sec	+	21.1	0.0021	790158:790234
BTB_c05490	tRNA-Sec	+	21.0	0.0022	531268:531342
BTB_c07960	tRNA-Sec	+	21.0	0.0022	789171:789245
BTB_c08010	tRNA-Sec	+	20.8	0.0025	789634:789707
BTB_c08080	tRNA-Sec	+	20.7	0.0026	790322:790397
BTB_c08130	tRNA-Sec	+	20.4	0.0032	790729:790800
BTB_c05400	tRNA-Sec	+	20.4	0.0032	530419:530493
BTB_c08110	tRNA-Sec	+	20.1	0.0038	790568:790644
BTB_c00090	tRNA-Sec	+	20.0	0.004	10777:10852
seq250	tRNA-Sec	-	19.7	0.0047	5128522:5128674
BTB_c05480	tRNA-Sec	+	19.5	0.0056	531129:531204
BTB_c05390	tRNA-Sec	+	19.0	0.0073	530310:530401
BTB_c02800	tRNA-Sec	+	18.9	0.0077	243427:243512
seq6	Bacteria_large_SRP	+	242.0	6.8e-47	24793:25179
BTB_c03160	SSU_rRNA_eukarya	+	742.1	4.8e-225	282956:284495
BTB_c03340	SSU_rRNA_eukarya	+	742.1	4.8e-225	305524:307063
BTB_c00070	SSU_rRNA_eukarya	+	741.9	5.7e-225	9017:10556
BTB_c00980	SSU_rRNA_eukarya	+	741.2	8.8e-225	87752:89291
BTB_c02840	SSU_rRNA_eukarya	+	739.8	2.3e-224	243920:245459
BTB_c07910	SSU_rRNA_eukarya	+	738.2	7.4e-224	784157:785696
seq30	FsrA	-	17.6	0.0064	202156:202443
seq201	ncr1175	+	25.5	0.00042	1530296:1530355
seq201	ncr1175	-	20.1	0.0068	1530296:1530355
BTB_c03070	LSU_rRNA_archaea	+	1998.6	0	268730:271649
BTB_c03120	LSU_rRNA_archaea	+	1998.6	0	275963:278882
BTB_c03350	LSU_rRNA_archaea	+	1998.6	0	307217:310136
BTB_c00990	LSU_rRNA_archaea	+	1997.8	0	89445:92364
BTB_c07920	LSU_rRNA_archaea	+	1990.6	0	785850:788769
BTB_c00990	LSU_rRNA_bacteria	+	3005.2	0	89445:92364
BTB_c03070	LSU_rRNA_bacteria	+	3004.0	0	268730:271649
BTB_c03120	LSU_rRNA_bacteria	+	3004.0	0	275963:278882
BTB_c03350	LSU_rRNA_bacteria	+	3004.0	0	307217:310136
BTB_c07920	LSU_rRNA_bacteria	+	2999.0	0	785850:788769
BTB_c07920	LSU_rRNA_eukarya	+	1298.7	0	785850:788769
BTB_c00990	LSU_rRNA_eukarya	+	1296.2	0	89445:92364
BTB_c03070	LSU_rRNA_eukarya	+	1296.0	0	268730:271649
BTB_c03120	LSU_rRNA_eukarya	+	1296.0	0	275963:278882
BTB_c03350	LSU_rRNA_eukarya	+	1296.0	0	307217:310136
seq201	rliE	-	21.6	0.0054	1530296:1530355
seq100	rli61	+	27.8	0.00012	1215721:1215858

Tabela 8: Relação entre as amostras

	Amostra 1	Amostra 2
Sequências extraídas	157	157
Candidatos Significativos	131	177
Candidatos Não Significativos	304	281
Candidatos não significativos resgatados	21	25
Total de possíveis candidatos	152	202

Tabela 9: Sequências não significativas resgatadas em novo limite de inclusão - Amostra 1

Nome	Família	Posição	Score	E-value	Coordenadas
seq29	TPP	+	17.6	0.011	429121:429222
seq121	PyrR	-	22.4	0.015	4477747:4477866
seq9	PyrR	+	22.4	0.015	111654:111807
seq97	PyrR	-	20.5	0.032	3663349:3663474
seq13	PyrR	+	18.8	0.064	182860:182988
seq70	L10_leader	+	17.7	0.029	1416921:1417173
seq76	L10_leader	+	17.5	0.031	1530292:1530418
seq88	RsaE	+	17.4	0.033	2928641:2929030
BTB_c07940	tRNA-Sec	+	17.1	0.017	788992:789067
seq70	PyrG_leader	+	20.2	0.046	1416921:1417173
seq76	PyrG_leader	+	18.8	0.086	1530292:1530418
seq121	ncr1015	-	18.6	0.076	4477747:4477866
seq88	ncr1175	+	17.4	0.02	2928641:2929030
seq160	ncr1241	-	17.0	0.046	5366140:5366383
seq44	Spy392987	-	17.8	0.031	741019:741078
BTB_c22130	RsaA	+	18.5	0.019	2115831:2116562
seq121	SpF41_sRNA	-	20.1	0.017	4477747:4477866
seq117	tsr26	+	18.5	0.05	4400739:4400496
seq88	tsr31	-	18.6	0.03	2928641:2929030
BTB_c44350	tsr33	+	18.4	0.14	4296553:4297680
seq75	psm_mec_RNA	-	18.0	0.097	1496274:1496355

Tabela 10: Sequências não significativas resgatadas em novo limite de inclusão - Amostra 2

Nome	Família	Posição	Score	E-value	Coordenadas
seq251	TPP	+	17.7	0.014	290777:290957

seq55	PyrR	+	23.1	0.015	337326:337443
seq222	PyrR	-	23.0	0.015	3925531:3925590
seq223	PyrR	-	21.8	0.025	1574120:1574300
seq220	PyrR	+	20.5	0.043	3787227:3787398
seq220	PyrR	-	17.0	0.18	3787227:3787398
seq222	L10_leader	-	20.1	0.013	3925531:3925590
seq55	rli38	+	20.8	0.037	337326:337443
seq30	RsaE	-	20.3	0.01	202176:202293
seq213	RsaE	+	17.4	0.045	2928645:2929014
BTB_c00680	tRNA-Sec	+	18.5	0.01	64015:64088
BTB_c05410	tRNA-Sec	+	17.1	0.023	530498:530573
seq104	PyrG_leader	+	20.2	0.062	1416849:1417218
seq203	ncr1015	-	19.0	0.086	1550689:1551204
seq236	ncr1015	-	17.0	0.22	4477749:4477857
seq213	ncr1175	+	17.4	0.027	2928645:2929014
seq220	ncr1575	-	22.3	0.011	3787218:3787407
seq223	ncr982	-	17.6	0.15	1574147:1574273
seq201	rliE	+	19.6	0.014	1530296:1530355
seq107	rli52	+	18.4	0.071	1492310:1492366
seq94	tracrRNA	+	17.9	0.043	1074994:1075329
seq208	sau-59	-	18.2	0.14	2307571:2307841
seq250	SpF41_sRNA	-	21.3	0.013	5128505:5128705
seq249	SpF41_sRNA	-	18.9	0.04	5109009:5109522
seq213	tsr31	-	18.6	0.04	2928645:2929014