

UNIVERSIDADE FEDERAL DO PARANÁ

LEANDRO FERREIRA MORENO

**PREDIÇÃO E ANÁLISE DA EXPRESSÃO DE RNAS NÃO
CODIFICADORES COM FUNÇÃO REGULATÓRIA PRESENTES NA
BACTÉRIA *HERBASPIRILLUM SEROPEDICAE* SmR1**

CURITIBA
2013

LEANDRO FERREIRA MORENO

**PREDIÇÃO E ANÁLISE DA EXPRESSÃO DE RNAs NÃO
CODIFICADORES COM FUNÇÃO REGULATÓRIA PRESENTES NA
BACTÉRIA *HERBASPIRILLUM SEROPEDICAE* SmR1**

**Dissertação apresentada como requisito
parcial à obtenção do Grau de Mestre
em Bioquímica, Curso de Pós-Graduação
em Ciências-Bioquímica, Setor de Ciências
Biológicas, Universidade Federal do
Paraná.**

**Orientadora: Profa. Maria Berenice R. Steffens
Coorientador: Dr. Marcelo Müller dos Santos**

CURITIBA
2013

TERMO DE APROVAÇÃO



PREDIÇÃO E ANÁLISE DE EXPRESSÃO DE RNAS NÃO CODIFICADORES COM FUNÇÃO REGULATÓRIA PRESENTES NA BACTÉRIA *HERBASPIRILLUM SEROPEDICAR* SMR1

Dissertação aprovada como requisito parcial para a obtenção do título de Mestre em Ciências-Bioquímica, no Programa de Pós-Graduação em Ciências – Bioquímica, Setor de Ciências Biológicas da Universidade Federal do Paraná, pela seguinte banca Examinadora:

Prof.^a Dr.^a Maria Berenice R. Steffens (Orientadora)
Departamento de Bioquímica e Biologia Molecular – UFPR

Dr. Helisson Faoro
Departamento de Bioquímica e Biologia Molecular – UFPR

Prof. Dr. Alexandre Rossi Paschoal
Universidade Tecnológica Federal do Paraná – UTFPR

Curitiba, Maio de 2013.

AGRADECIMENTOS

À Deus pelo suporte espiritual.

Agradeço à minha orientadora, Professora Dra. Maria Berenice Reynaud Steffens, pela confiança, troca de conhecimentos, amizade, carinho e ótima convivência nesses últimos 5 anos. Muito Obrigado.

Ao meu coorientador Professor Dr. Marcelo Müller dos Santos pelas saudáveis discussões sobre este trabalho.

À minha noiva Juliana, por me manter feliz em todos os momentos, dentro e fora do laboratório.

À minha família, por todo apoio e incentivo.

Aos Professores Dr. Fábio de Oliveira Pedrosa e Dr. Emanuel Maltempi de Souza, pelas discussões, ideias e pelo tempo disponibilizado.

À Professora Jeroniza Nunes Marchaukoski, pela oportunidade de iniciar as pesquisas com ncRNAs.

Aos amigos do Laboratório de Bioinformática do Departamento de Bioquímica da UFPR, Rodrigo, Vinicius, Felipe e Helisson pelas ideias, boa convivência, incentivo e companheirismo.

Ao professor Claudio Valverde da Universidade Nacional de Quilmes, pelo suporte e ajuda nas análises realizadas neste trabalho.

A todos meus amigos de mestrado por me lembrarem de que na vida nem tudo é científico.

À coordenação de Pós-Graduação em Bioquímica e a todas as Instituições Financiadoras, particularmente ao INCT da Fixação Biológica de Nitrogênio pelo apoio financeiro e a CAPES/REUNI e CNPq pela bolsa de estudos.

Muito Obrigado!

RESUMO

Herbaspirillum seropedicae estirpe SmR1 é uma bactéria endofítica capaz de fixar nitrogênio e promover o crescimento de importantes culturas agrícolas. Seu genoma foi completamente sequenciado e anotado pelo Programa Genoma do Estado do Paraná (GENOPAR Consortium (www.genopar.org)). Esta bactéria tem um único cromossomo circular de 5.513.887 pares de base com 4.735 ORFs anotadas, as quais representam 88,3% do genoma. Em bactérias, RNAs não codificadores com função regulatória (ncRNAs) podem modular várias respostas fisiológicas e atuar por diferentes mecanismos, como pareamento de bases de RNA-RNA e interações RNA-proteína. Tecnologias de sequenciamento *High-throughput*, como por exemplo, a plataforma SOLiD, estão permitindo a identificação em larga escala de ncRNAs, revelando a existência de vários transcritos não-codificadores e indicando que a quantidade de ncRNAs reguladores pode ser maior do que se pensava anteriormente. Tradicionalmente, abordagens *in silico* para a identificação dessas moléculas envolvem a associação de sequência de promotor fator sigma70 com sequências de terminador Rho-independentes, e / ou conservação de estruturas primárias e secundárias de RNA. O objetivo deste trabalho foi identificar e avaliar a expressão de ncRNAs presentes em *H. Seropedicae* SmR1. Para isso, o genoma completo foi pesquisado com as ferramentas de bioinformática Gsalgorithm e nocoRNAC, que foram usadas para identificar regiões do genoma flanqueadas por sequências de promotor e sequências de terminador Rho- independentes candidatas a codificar ncRNAs. Adicionalmente, a ferramenta *Cufflinks* foi utilizadas para localizar regiões do genoma com consideráveis níveis de transcrição e livres de ORFS. Para avaliar a expressão dos ncRNAs, o transcriptoma de *H. seropedicae* SMR1 cultivada em três diferentes condições foi determinado por RNA-Seq utilizando a plataforma de sequenciamento SOLiD. Um total de 98 ncRNAs foram confirmados no conjunto de dados do transcriptoma. A comparação dos ncRNAs com os bancos de dados RFAM e RIBEX revelaram que apenas oito transcritos identificados nesse estudo já haviam sido descritos em outras espécies de bactérias, descando os seguintes: 4.5S RNA, 6S RNA (SsrS), Intron_gpl, tmRNA, and TPP Riboswitch. A função de 90 novos ncRNAs serão investigados *in vivo*.

Palavras-chave: *Herbaspirillum seropedicae* SmR1, ncRNA, sRNA, cis-RNA, Riboswitch.

ABSTRACT

Herbaspirillum seropedicae strain SmR1 is an endophytic bacterium capable of fixing nitrogen and promoting growth of some important agricultural crops. Its complete genome was sequenced and annotated by the Paraná State Genome Program (Genopar Consortium (www.genopar.org)). This bacterium has a single circular chromosome of 5,513,887 base pairs with 4,735 annotated open reading frames, which represent 88.3% of the genome. Bacterial regulatory non-coding RNAs (ncRNA) can modulate various physiological responses acting by different mechanisms such as RNA–RNA base-pairing and RNA–protein interactions. High-throughput sequencing technology such as the SOLiD platform has allowed genome scale identification of ncRNAs, revealing the existence of several noncoding transcripts and indicating that the amount of regulatory ncRNAs can be higher than previously thought. Several computational tools have been designed to identify ncRNA in bacterial genome sequences. RNAz, sRNAFinder, sRNAPredict, sRNAscanner and Infernal have been used to identify an extensive range of new ncRNA molecules. Traditional *in silico* approaches for identifying ncRNAs involve the association of sigma70 promoter sequences with Rho-independent terminator sequences and/or conservation of primary and secondary RNA structures. The goal of this work was to identify and validate ncRNAs of *H. seropedicae* SmR1. To accomplish this, the complete genome was screened with the bioinformatics tools GSalgorithm and nocoRNAC to identify genomic regions flanked by promoter sequences and terminator Rho-independent sequences likely to encode ncRNAs. Additionally the tool Cufflinks figured out transcript regions likely to be ncRNAs. To address the expression of the ncRNAs, the transcriptome of *H. seropedicae* Smr1 grown under three distinct conditions was determined by RNA-Seq in a SOLiD sequencing platform. A total of 98 ncRNAs were confirmed in the transcriptome data set. Comparison of these ncRNAs with the RFAM database revealed that only eight transcripts had been previously described in other bacterial species, such as the 4.5S RNA, 6S RNA (SsrS), Intron_gpl, tmRNA, and TPP Riboswitch. The function of the 90 novel ncRNAs will be investigated *in vivo*.

Key-words: *Herbaspirillum seropedicae* SmR1, ncRNA, sRNA, cis-RNA, Riboswitch.

LISTA DE ABRAVIATURAS

bp – Pares de base

IgR (Regiões Intergênicas) – Região do genoma que não contém ORFs

gbK – Extensão do arquivo que armazena sequência genética

mRNA – RNA mensageiro

NCBI - *National Center for Biotechnology Information*

ncRNA – RNA não codificador de proteína

ORF – “*open reading frame*” ou sequência codificadora de proteína

RNA – Ácido Ribonucléico

RPKM – *reads per kilobase of transcript per million mapped reads*

rRNA – RNA ribossomal

sRNA – RNA não codificador de proteína com função regulatória

tRNA – RNA transportador

UTR (Untranslated Region) – Sequência de bases não traduzidas em um mRNA

WEB (World Wide Web) – Rede de internet de alcance mundial

SUMÁRIO

1 INTRODUÇÃO	10
2 OBJETIVO	12
2.1 Objetivo Geral	12
2.2 Objetivos Específicos	12
3 REVISÃO BIBLIOGRÁFICA	13
3.1 RNAs não codificadores bacterianos.....	13
3.1.1 trans-sRNAs	14
3.1.2 cis-sRNAs.....	16
3.1.3 <i>Riboswitches</i>	18
3.2 Abordagens para identificação de ncRNAs	20
3.2.1 Métodos Computacionais de predição de ncRNAs	21
3.2.2 Métodos Experimentais de predição de ncRNAs	21
3.2.3 Banco de dados de ncRNAs	22
3.2.4 Predição de alvo da ação de ncRNAs	23
3.3 <i>Herbaspirillum seropedicae</i>	24
4. MATERIAIS E MÉTODOS	26
4.1 Conjunto de dados	26
4.1.1 Genoma.....	26
4.1.2 Sequências de ncRNA	26
4.1.3 Dados de RNA-seq.....	26
4.2 Sistema operacional	27
4.3 Linguagens de Programação	27
4.4 Ferramentas de Bioinformática	27
4.4.1 Artemis	27
4.4.2 CuffLinks	28
4.4.3 cmsearch.....	28
4.4.4 GSalgorithm	28
4.4.5 (non-coding RNA characterization) nocoRNAC	29
4.4.6 gsmapper	30
4.4.7 SamTools	30
4.5 Hardware.....	31
4.6 Transcriptoma	31
4.6.1 Mapeamento de Leituras de RNA-seq	32
4.6.2 Análise do transcriptoma	33

5. RESULTADOS	35
5.1 Identificação de ncRNAs utilizando os programas GSalgorithm e nocoRNAc	35
5.2 Identificação de ncRNAs utilizando dados de RNA-seq.....	40
5.3 ncRNAs confirmados experimentalmente em <i>H. seropedicae</i> SmR1.....	42
5.4 Conservação dos ncRNAs	49
5.5 Análise de dois prováveis <i>Riboswitches</i> : RLE0100 e RLE0038	51
5.7 Alvos dos ncRNAs de <i>H. seropedicae</i> SmR1 confirmados experimentalmente	52
6 DISCUSSÃO	55
7 CONCLUSÕES	59
REFERÊNCIAS BIBLIOGRÁFICAS	60
APÊNDICES	69

1 INTRODUÇÃO

Herbaspirillum seropedicae é uma bactéria diazotrófica endofítica, pertencente ao grupo das beta-proteobactérias, capaz de fixar nitrogênio e promover o crescimento de gramíneas economicamente importantes, tais como trigo, milho e arroz. *H. seropedicae* aparece associada também com plantas tropicais como bananeira e abacaxizeiro. Essa bactéria possui um único cromossomo circular com 5.513.887 pares de nucleotídeos que estima-se conter 4.735 *open reading frames* (ORF). Assim como em outros organismos, expressa um conjunto de RNAs não codificadores, com funções regulatórias (ncRNA), cuja caracterização pode fornecer informações fundamentais sobre a fisiologia desse organismo. RNAs não codificadores bacterianos podem modular diversas respostas fisiológicas através de diferentes mecanismos de ação, como o pareamento de base mRNA-ncRNA e proteína-ncRNA. A classe mais estudada de ncRNAs é conhecida como sRNA (*small RNA*) e sua transcrição ocorre em orientação *trans* ao seu alvo. Por outro lado, ncRNAs que são transcritos no sentido *cis* são denominados de ncRNAs anti-senso. Outras moléculas de ncRNA recebem o nome de *Riboswitch* e estão frequentemente localizados na região 5' UTR de um mRNA, podendo promover a regulação transcricional através de sua interação com uma molécula ligante. Os ncRNAs podem ser preditos por meio de ferramentas computacionais que buscam características importantes presentes nessa molécula, tais como a presença de promotor, tipicamente dependente de fator sigma70, e terminador Rho-Independente em regiões intergênicas ou em regiões antisenso a ORFs. Novas abordagens, como o sequenciamento de RNA em larga escala (RNA-seq), tem sido amplamente utilizadas na predição de ncRNAs, pois são capazes de sugerir a presença dessas moléculas com grande eficiência. Atualmente, diversos estudos relatam a existência de sRNAs em genomas de bactérias Gram-negativas e Gram-positivas. Porém, poucas análises, utilizando o gênero *Burkholderia*, descrevem a presença dessas moléculas na classe das Betaproteobacterias. Pouco se conhece sobre os ncRNAs em *H. seropedicae* estirpe SmR1. O trabalho de Willemann (2010), utilizando uma ferramenta de busca de sRNAs, sugeriu a presença de 57 genes candidatos no genoma dessa bactéria, porém nenhuma confirmação experimental envolvendo estes candidatos foi realizada até o momento. A proposta deste trabalho é avançar

na prospecção de ncRNAs em *H. seropedicae* e confirmar sua presença através da análise transcricional utilizando novas tecnologias de sequenciamento.

2 OBJETIVO

2.1 Objetivo Geral

Predizer e avaliar a expressão de RNAs não codificadores com função regulatória na bactéria *Herbaspirillum seropedicae* SmR1 utilizando técnicas computacionais e dados de sequenciamento de RNAs expressos (RNA-seq).

2.2 Objetivos Específicos

- a) Utilizar os programas de bioinformática *GSAlgorithm* e *nocoRNAc* para a identificação de possíveis genes de ncRNAs presentes no genoma de *H. seropedicae* SmR1.
- b) Mapear no genoma de *H. seropedicae* SmR1 leituras derivadas de experimentos de RNA-seq.
- c) Analisar o perfil de expressão dos ncRNAs candidatos.
- d) Utilizar a ferramenta de bioinformática *cufflinks* e dados de RNA-seq para identificar regiões no genoma com significativos níveis de transcrição, que não correspondam a ORFs e que possam ser caracterizadas como ncRNAs.
- e) Sugerir a função dos ncRNAs confirmados através de comparação de estrutura primária e secundária com sequências depositadas no banco de dados RFAM, RIBEX e sRNAMap.
- f) Realizar uma análise comparativa dos ncRNAs identificados em *H. seropedicae* SmR1 com o genoma parcial da bactéria *Herbaspirillum rubrisubalbicans* M1.
- g) Ampliar o conhecimento acerca dos ncRNAs de Betaproteobactérias

3 REVISÃO BIBLIOGRÁFICA

3.1 RNAs não codificadores bacterianos

RNAs não codificadores (ncRNAs) são expressos em organismos nos três domínios da vida, *Bacteria*, *Archaea* e *Eukarya*. Até o ano de 2012, em *Escherichia coli*, o microrganismo onde o maior número ncRNAs foi descrito, aproximadamente 80 dessas moléculas foram identificadas e validadas. Entretanto, análises computacionais indicam que esse número seja muito maior (RAGHAVAN; GROISMAN; OCHMAN, 2011; WATERS; STORZ, 2009). Diversos estudos relatam a existência de ncRNAs em genomas de bactérias Gram-negativas e Gram-positivas (SHARMA; VOGEL, 2009). Porém, poucas análises, utilizando o gênero *Burkholderia*, descrevem a presença dessas moléculas na classe das Betaproteobacterias (COENYE et al., 2007; YODER-HIMES et al., 2009). Em eucariotos o número de ncRNAs descritos para 140 espécies é superior a 15 mil, como indica um dos principais repositórios online dessas moléculas (KOZOMARA; GRIFFITHS-JONES, 2011).

Os primeiros ncRNA identificados em uma bactéria datam da década de 70. Essas moléculas, nomeadas então de 6S e 4.5S RNAs, foram localizadas utilizando eletroforese de gel de poliacrilamida (WASSARMAN; ZHANG; STORZ, 1999). Desde então, centenas de genes candidatos de ncRNAs com função regulatória tem sido preditos. Os ncRNAs bacterianos possuem uma grande diversidade de tamanho podendo variar de ~50 até ~600 nucleotídeos (PICHON; FELDEN, 2008). Outras características relacionadas aos ncRNAs bacterianos incluem a presença de sequência de promotor, sequência de terminadores Rho-Independente e a conservação na sequência primária e secundária quando comparados com ncRNAs de espécies filogeneticamente relacionadas. (ARGAMAN et al., 2001; CHEN et al., 2002).

A maioria dos ncRNAs com função já descrita, regulam a expressão gênica se ligando a mRNAs ou a proteínas (VOGEL; WAGNER, 2007). Os ncRNAs em bactérias agem de diferentes maneiras modulando uma ampla gama de respostas fisiológicas, tais como replicação de DNA e divisão celular (*dicF*), regulação transcricional (6S RNA), processamento de RNA (RNase P ou *rnpB*), estabilidade

de mRNA (*spotou42*), estabilidade de proteína (*tmRNA*) e transporte (*4.5S ffs*), resposta ao estresse oxidativo (*oxyS*), controle do número de cópias de plasmídeo (RNAI e RNAIII), e armazenamento de carbono (*csrBC*) (ULVE et al., 2007; WATERS; STORZ, 2009). Como reguladores de mRNAs, os ncRNAs podem modular positivamente ou negativamente a tradução de um gene (LEASE; BELFORT, 2000). Este efeito depende da exata localização onde ocorrerá a interação e da alteração na estrutura secundária do mRNA alvo (EGGENHOFER et al., 2011). A interação entre o ncRNA e o mRNA pode levar à repressão traducional ou ao silenciamento irreversível através da degradação da molécula alvo, resultando na redução do produto gênico. De forma contrária, essa interação pode expor a sequência Shine-Dalgarno facilitando o acesso do ribossomo ao mRNA aumentando a eficiência no processo de tradução (LEASE; BELFORT, 2000; MAJDALANI et al., 2001). Frequentemente, essas interações são estabilizadas pela proteína Hfq, uma chaperona conservada em muitas espécies bacterianas (LIU; CAMILLI, 2010). Outra forma de regulação induzida pelos ncRNAs envolve a alteração da atividade de proteínas. Nesse caso, as proteínas se tornam alvos diretos dos ncRNAs, como ocorre nos sRNAs 6S e CrsB em *E. coli* (BABITZKE; ROMEO, 2007; STORZ; OPDYKE; ZHANG, 2004; WILLKOMM; HARTMANN, 2005).

Atualmente não existe uma nomenclatura oficial organizando as diferentes classes de ncRNAs. Recentes publicações sugerem a organização dos ncRNAs em cis-sRNA, trans-sRNA e *Riboswitches*. Esta classificação é baseada na localização e na forma de ação desta molécula sobre seu alvo. (D'ALIA et al., 2010; HOUSELEY et al., 2008; PICHON; FELDEN, 2008; WATERS; STORZ, 2009).

3.1.1 trans-sRNAs

Os trans-sRNAs são transcritos em regiões genômicas distantes de seu mRNA alvo e geralmente compartilham pouca complementariedade de bases com o mesmo (PICHON; FELDEN, 2008). Comumente, a interação entre estes ncRNAs e seus alvos ocorrem por pequenos e imperfeitos pareamentos de bases, possibilitando a existência de múltiplos alvos. Muitos trans-ncRNAs se ligam à região do sítio de ligação do ribossomo (RBS) do mRNA alvo bloqueando o processo de tradução. Adicionalmente, a ligação do trans-sRNA pode aumentar ou diminuir a estabilidade do mRNA (WATERS; STORZ, 2009). Para a estabilização da interação

deste ncRNA com seu alvo, a chaperona Hfq pode ser necessária (AIBA, 2007; PAPPENFORT; VOGEL, 2009; ZHANG et al., 2003). Esta proteína promove o anelamento entre todos os sRNAs dependentes de Hfq e seus respectivos alvos. Essa propriedade acontece em decorrência da alteração da estrutura secundária dos RNAs em resposta à ligação da Hfq que pode expor o sítio de pareamento de bases. É possível ainda que a Hfq aumente a concentração dos RNAs envolvidos no pareamento de bases ligando-se simultaneamente no ncRNA e em seu alvo. (AIBA, 2007; STORZ et al., 2004). Os mecanismos descritos estão representados na FIGURA 1. Outro papel importante realizado pela proteína Hfq é a proteção do mRNA alvo ou do ncRNA contra a digestão por RNase E. A função de proteção é decorrente da competição entre a proteína Hfq e a RNase E pelo mesmo sítio de ligação (MOLL et al., 2003; STORZ et al., 2004; ZHANG et al., 2003).

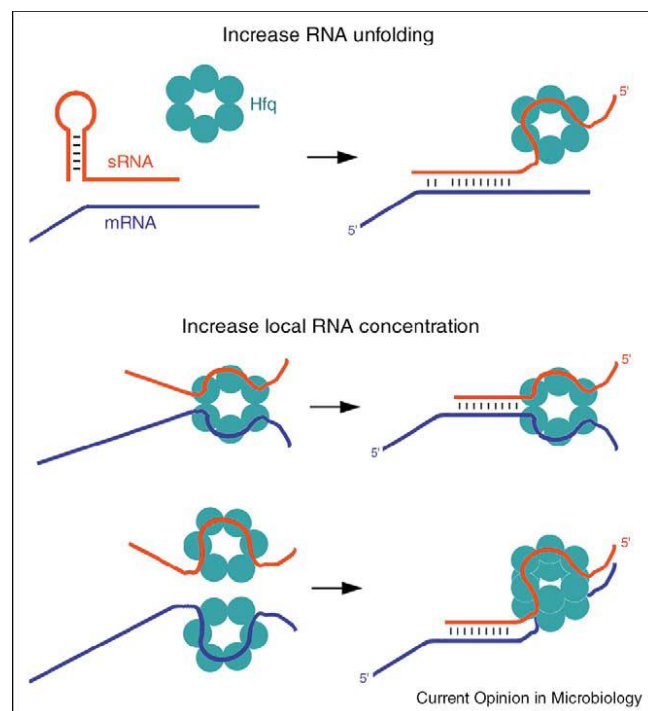


FIGURA 1 - A proteína Hfq pode facilitar o pareamento ncRNA-mRNA. Hfq (hexâmero azul) pode promover a mudança na estrutura dos RNAs ou pode ligar-se simultaneamente ao ncRNA (vermelho) e ao mRNA alvo (azul) formando um complexo ribonucleoprotéico. FONTE: (STORZ et al., 2004).

Os trans-sRNAs podem regular a atividade de proteínas sequestrando-as através dos sítios de ligação de RNA presente nessas moléculas. A associação do sRNA com a proteína pode inibir, modificar ou ativar a atividade proteica (STORZ; VOGEL; WASSARMAN, 2011).

3.1.2 cis-sRNAs

Os ncRNAs classificados como cis-sRNAs são transcritos antisenso ao seu mRNA alvo, regulando-o através de um perfeito pareamento de bases (GEORG; HESS, 2011). O mecanismo biológico desta regulação ainda não está completamente elucidado, mas acredita-se que essa interação altere a estrutura secundária das moléculas que estão interagindo, influenciando na estabilidade e na meia vida desses RNAs (LASA; TOLEDO-ARANA; GINGERAS, 2012). A ligação do cis-sRNAs com o mRNA produz uma molécula de RNA de dupla fita que pode ser degradada pela enzima RNase III produzindo fragmentos de aproximadamente 20 pares de bases (LASA et al., 2011). O reconhecimento do mRNA alvo ocorre de forma rápida e com alta afinidade (STORZ et al., 2011). Estima-se que a quantidade de regiões transcritas antisenso a ORFs seja muito grande em genomas bacterianos. Em *E. coli*, por exemplo, foram encontrados aproximadamente quatro mil transcritos antisenso a ORFs, o que pode indicar a presença de muitas moléculas de cis-sRNAs ainda não descritas (SELINGER et al., 2000). O gene destes ncRNAs podem estar em oposição a sequências 5' ou 3', no meio ou por todo o gene de mRNA alvo (THOMASON; STORZ, 2010a). Com relação à função, estudos apontam para um envolvimento na inibição da tradução do mRNA alvo, porém, existem relatos sobre a ativação da expressão de genes por cis-sRNAs (WAGNER; ALTUVIA; ROMBY, 2002). Na FIGURA 2 estão esquematiza as diferentes formas de ação de um cis-sRNA sobre um mRNA.

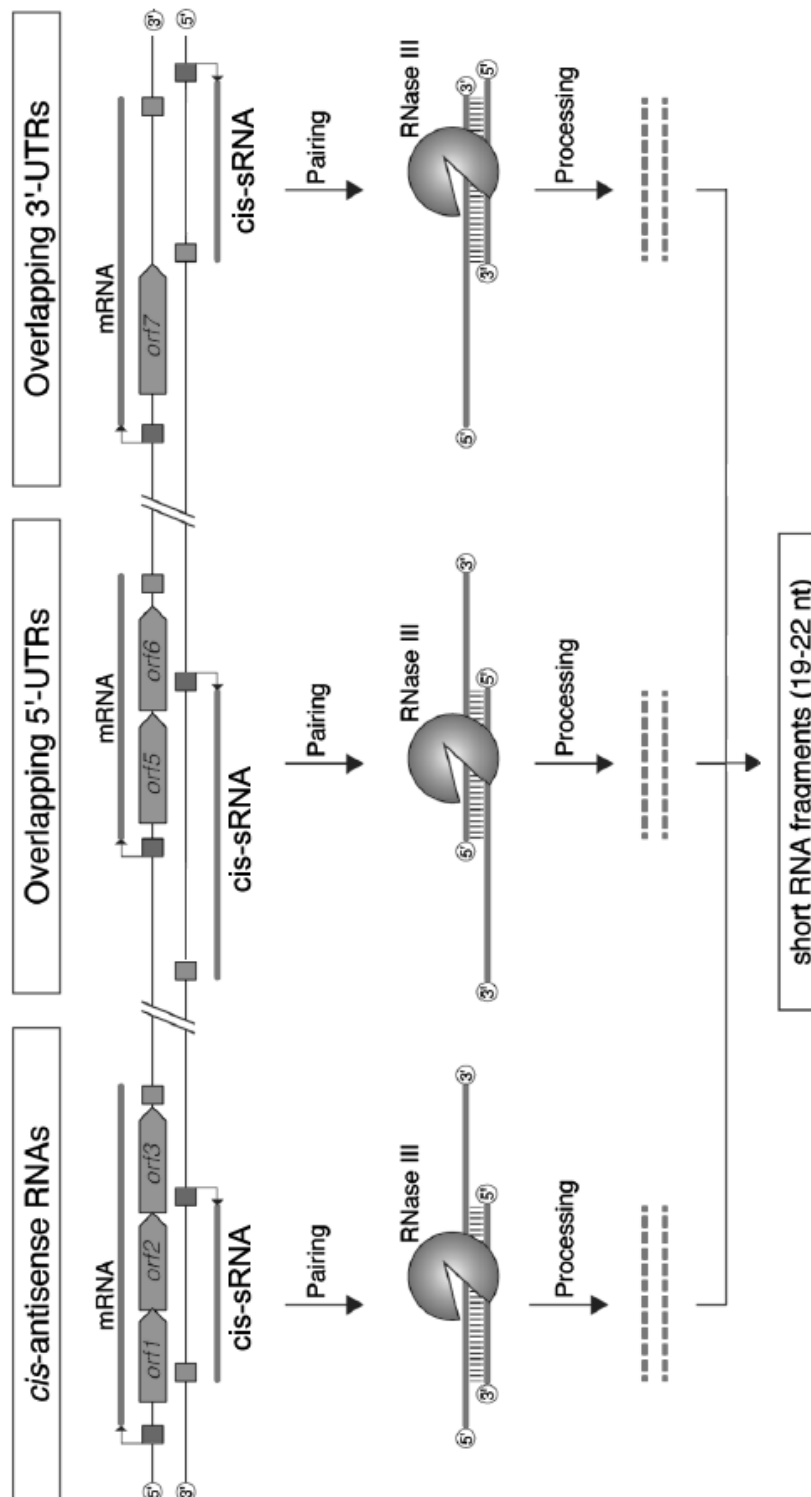


FIGURA 2 - Mecanismo de ação de uma molécula de cis-sRNA e posterior processamento pela endoribonuclease RNase III. O ncRNA pode estar antisense a seqüências 5' ou 3', no meio ou por todo o gene de mRNA alvo. O processamento realizado pela RNase III produz fragmentos de aproximadamente 20 pares de bases. FONTE: Adaptado de (LASA et al., 2012)

3.1.3 *Riboswitches*

Riboswitches são moléculas de RNA, localizadas na região 5' UTR de um mRNA, capaz de controlar a expressão gênica a nível de transcrição e de tradução (BREAKER, 2012). Os *Riboswitches* se ligam a moléculas denominadas de ligantes, que podem ser metabólitos ou íons metálicos. Esses elementos regulatórios são capazes de responder a mudanças ambientais, como por exemplo, a alteração de temperatura (SMITH et al., 2010). A ligação e o reconhecimento do ligante ou o estímulo ambiental acontecem em uma região conhecida como sensor, e provocam alterações na estrutura secundária do *Riboswitch* desencadeando o processo de regulação (JOHANSSON, 2009).

Acredita-se que o *Riboswitch* possua um grampo terminador intrínseco ou um antiRBS para regulação transcricional ou traducional, respectivamente (BARRICK; BREAKER, 2007). No que se refere à regulação transcricional, o *Riboswitch* liga ou desliga moléculas em seu sensor, permitindo que um grampo terminador intrínseco seja exposto ou ocultado levando a expressão ou inativação de um gene (WACHSMUTH et al., 2012). No que concerne à regulação traducional, uma estrutura denominada de antiRBS, é complementar ao sítio de ligação ribossomal do mRNA, e quando formada, impede que o ribossomo inicie o processo de tradução (TOPP; GALLIVAN, 2010). As FIGURAS 3 e 4 apresentam a regulação transcricional e traducional respectivamente.

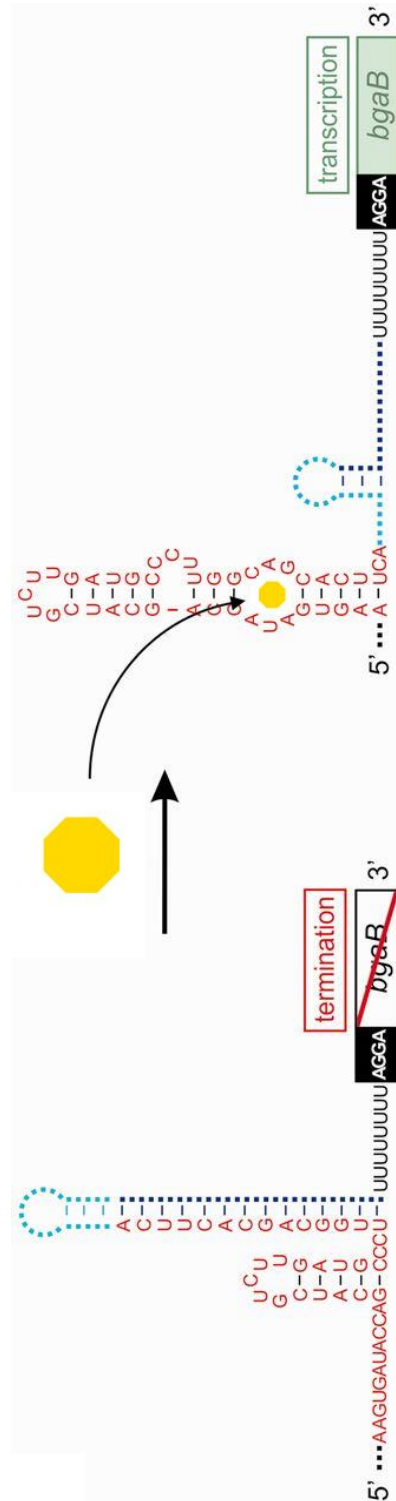


FIGURA 3 - Mecanismo de regulação transcricional realizado por um Riboswitch.

A sequência em vermelho representa a região do sensor. O pontilhado azul esquematiza o grampo terminador intrínseco do Riboswitch seguido por uma cauda poli-U. A sequência Shine-Dalgarno e o gene *bgaB* estão localizados a jusante do Riboswitch. Com a ligação do metabólito (amarelo) ao Riboswitch, a formação do grampo terminador é inibida resultando na transcrição do gene. Na ausência do ligante, a estrutura do grampo terminador se forma. FONTE: Adaptado de (WACHSMUTH et al., 2012)

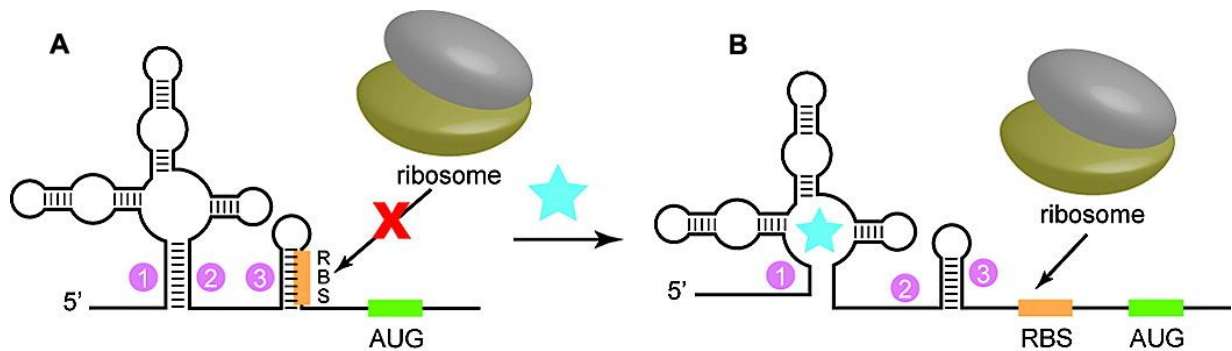


FIGURA 4 - Mecanismo de regulação traducional realizado por um Riboswitch. Estão numeradas as estruturas: 1) mRNA, 2) Riboswitch e 3) grampo antiRBS. O RBS do mRNA está pareado com o grampo antiRBS do Riboswitch impedindo o acesso do ribossomo. A ligação de um metabólito (Estrela azul) ao Riboswitch altera sua estrutura secundária deixando o grampo antiRBS distante do RBS do mRNA. Isso permite que o ribossomo inicie o processo de tradução. FONTE: Adaptado de (TOPP; GALLIVAN, 2010).

Devido à sua localização, os *Riboswitches* são sintetizados anteriormente ao mRNA, possibilitando que a resposta à ligação do metabólito aconteça antes da transcrição total do mRNA. Dessa forma, o processo de regulação é biologicamente viável, pois evita gastos energéticos na produção de mRNAs inteiros (BREAKER, 2011). Esses ncRNAs estão envolvidos em diversos processos celulares, como processamento de RNA, estabilidade de mRNA (EDWARDS, 2010; JOHANSSON, 2009).

3.2 Abordagens para identificação de ncRNAs

A procura por ncRNAs em genomas bacterianos tem sido realizada utilizando métodos computacionais, métodos experimentais ou combinando essas duas abordagens (ALBRECHT et al. 2010). A escolha do método compreende uma importante etapa do processo, isso porque o grupo de ncRNA que será encontrado varia de acordo com o método utilizado, por exemplo: estudos baseados somente na conservação de sequência do ncRNA e/ou em sua estrutura secundária podem falhar na identificação de ncRNAs espécie-específicos. Isso acontece, porque esse grupo de ncRNAs possuem características exclusivas, tornando esse tipo de método ineficaz. Da mesma forma, ncRNAs que são expressos sob diferentes condições

fisiológicas, podem não ser totalmente identificados em um estudo experimental que avalie sua transcrição em uma única condição de cultivo (ALTUVIA, 2007a).

3.2.1 Métodos Computacionais de predição de ncRNAs

Abordagens computacionais podem ser utilizadas para a predição de ncRNAs localizando, através de padrões de substituição, sequências de DNA conservadas e que estão presentes em regiões intergênicas de espécies relacionadas filogeneticamente (RIVAS; EDDY, 2001). Algumas ferramentas buscam identificar sinais de início e de término de transcrição distantes 50-600 pares de nucleotídeos, características observadas em ncRNAs previamente conhecidos por métodos de detecção direta (4.5S RNA, tmRNA, 6S RNA, RNaseP RNA, and Spot42 RNA) ou durante estudos da interação de ncRNAs com proteínas (OxyS RNA, Crp Tic RNA, CsrB RNA, e GcvB RNA). O foco é direcionado para sequências de promotores dependentes do fator σ^{70} e terminadores Rho-independentes (ALTUVIA, 2007b; ARGAMAN et al., 2001; RIVAS et al., 2001; WASSARMAN et al., 2001; WASSARMAN et al., 1999).

Atualmente, novos métodos computacionais mais eficientes possibilitam a análise de genomas bacterianos de espécies divergentes. Entre esses métodos destacam-se ferramentas que permitem o estudo de dados de experimentos de RNA-seq, utilização de modelos de estrutura de ncRNA para busca, entre outros (GRIFFITHS-JONES et al., 2003; TRAPNELL et al., 2012).

3.2.2 Métodos Experimentais de predição de ncRNAs

Entre as abordagens mais utilizadas, principalmente nos anos 90 e início dos anos 2000, destacam-se a detecção direta utilizando *microarray* e análises de *Nothern Blot*. No entanto, essas técnicas apresentam limitações como artefatos de hibridização cruzada e restrições no design das placas o que limita a detecção de genes não anotados previamente (MORTAZAVI et al., 2008). Devido a isso, ultimamente, essas abordagens têm sido mais úteis para corroboração de resultados do que identificação de novos ncRNA (WASSARMAN et al., 2001). A partir de 2004, novas metodologias de sequenciamento de DNA capazes de gerar um grande volume de dados de sequência em um menor tempo, com uma maior acurácia e

com um baixo custo tornaram-se disponíveis comercialmente (MARDIS, 2008; METZKER, 2010). Estas tecnologias baseiam-se no emprego de plataformas que permitem o sequenciamento paralelo e maciço de DNA e a produção de centenas de milhares de leituras de curto (50-100 pb) a médio (300 pb) tamanho.

Os experimentos utilizando as novas plataformas de sequenciamento são capazes de detectar o conjunto completo de transcritos, e sua quantidade, em uma célula crescendo em uma específica condição fisiológica. Esta abordagem é denominada de transcriptoma, e é essencial para a interpretação dos elementos funcionais de um genoma (WANG; GERSTEIN; SNYDER, 2009). Nela, uma população de RNA é convertida para uma biblioteca de fragmentos de cDNA ligados a adaptadores, que posteriormente são sequenciados. O sequenciamento do transcriptoma inteiro é nomeado de RNA-seq e é capaz de catalogar todas as espécies de RNAs, como mRNA, rRNA, tRNA, ncRNA entre outras (MORTAZAVI et al., 2008; VAN VLIET, 2010).

Estudos do transcriptoma total de diferentes organismos têm revelando a existências de transcritos até então desconhecidos, dentre estes, algumas classes de ncRNAs (LANGENBERGER et al., 2010).

3.2.3 Banco de dados de ncRNAs

Para lidar com o número crescente de ncRNAs sendo descritos, banco de dados que catalogam e armazenam informações como tamanho, localização e estrutura secundária dessa categoria de RNAs tem sido desenvolvido. Através de alinhamentos de sequências e modelos estatísticos, os bancos de dados permitem identificar similaridade entre um ncRNA de interesse e ncRNAs já descritos em outras espécies. Entre os bancos de dados mais utilizados para estudos de ncRNAs de bactérias destacam-se o RIBEX (ABREU-GOODGER; MERINO, 2005) por armazenar informações sobre riboswitches bacterianos, o RFAM (GARDNER; DAUB; et al., 2011) por organizar as moléculas de ncRNAs em famílias que representam um conjunto de sequências que compartilham um ancestral comum e o sRNAMap (HUANG et al., 2009) especializado em ncRNAs de bactérias gram-negativas. O trabalho de busca de informações em bancos de dados é facilitado com a função de busca online disponibilizada por essas aplicações, onde o usuário interessado pode identificar a qual classe seu candidato pertence sem a

necessidade de desenvolver algoritmos específicos. A quantidade de ncRNAs armazenada nos bancos de dados varia, como mostra a TABELA 1.

TABELA 1. Número de sRNA depositados em diferentes bancos de dados até novembro de 2012.

BANCO DE DADOS	NÚMERO DE SRNA
RFAM	252*
RIBEX	358
sRNAMap	331

* Total contabilizado sem repetição entre espécies.

Fonte: Autor (2012)

3.2.4 Predição de alvo da ação de ncRNAs

Grande parte dos ncRNAs identificados em bactérias interagem com mRNAs alterando a estabilidade e/ou influenciando o processo de tradução de seu alvo (EGGENHOFER et al., 2011). Abordagens experimentais como *knockouts* gênicos ou superexpressão do ncRNA de interesse permitem a identificação de uma parcela desses alvos, no entanto, exigem disponibilidade de tempo e grandes esforços (PICHON; FELDEN, 2008). Os ncRNAs em sua maioria não possuem alvos de ação identificados ou validados em laboratório até o momento. O pareamento de bases entre o sRNA e seu mRNA alvo acontece de forma imperfeita o que possibilita a existência de vários alvos sob influência de um mesmo sRNA dificultando a tarefa de predição (THOMASON; STORZ, 2010b). Para facilitar a localização de possíveis mRNAs alvos, algumas aplicações baseadas em servidores web tem sido desenvolvidas. Essas ferramentas permitem a identificação *in silico* de alvos de ncRNAs baseando-se na complementariedade de bases entre o ncRNA e seu alvo e em valores termodinâmicos decorrentes desta interação (BACKOFEN; HESS, 2010). Para avaliar a complementariedade de bases, métodos de alinhamento local, como por exemplo, o programa BLAST (ALTSCHUL et al., 1990), podem ser utilizados, porém, abordagens mais específicas para predição de alvos, como a ferramenta TargetRNA (TJADEN, 2008), atribuem um *score* ao pareamento de bases ao mesmo tempo em que calculam valores de energia livre mínima gerando um ranking com os melhores genes candidatos a alvos. O QUADRO 1 fornece informações sobre

diferentes ferramentas para predição de alvos de ncRNA e suas principais características.

QUADRO 1 - FERRAMENTAS PARA A PREDIÇÃO DE ALVOS DE ncRNA.

NOME	ABORDAGEM UTILIZADA	URL PARA ACESSO
TargetRNA	Pareamento de base/Energia livre mínima	http://snowwhite.wellesley.edu/targetRNA/
IntaRNA	Energia livre mínima/Acessibilidade da região de interação	http://www.bioinf.uni-freiburg.de/Software/
RNAup	Valores termodinâmicos da interação RNA-RNA	http://www.tbi.univie.ac.at/~ulim/RNAup/
RNAhybrid	Energia livre mínima	http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/
RNApredator	Valores termodinâmicos da interação RNA-RNA	http://rna.tbi.univie.ac.at/RNApredator

Fonte: Autor (2012)

3.3 *Herbaspirillum seropedicae*

O gênero *Herbaspirillum* pertence à subclasse β das proteobactérias e compreende 13 espécies entre elas *Herbaspirillum seropedica* (BALDANI et al., 1986; SCHMID, 2006). Estas bactérias são gram-negativas e tem um formato celular vibrióide ou helicoidal com um diâmetro de aproximadamente 0,7-07 μm . Apresentam um a três flagelos em ambos os pólos e por isso são bastante móveis (SCHMID, 2006). A espécie *Herbaspirillum seropedicae* é endofítica e fixa nitrogênio sob condições microaeróbicas (Baldani et al 1986). Relativamente à associação com plantas, a bactéria *H. seropedicae* é encontrada colonizando raízes, caules e folhas de espécies de gramíneas economicamente importantes (RONCATO-MACCARI et al., 2003). Esta propriedade faz do *H. seropedicae* um microorganismo com elevado potencial para ser utilizado como biofertilizante para culturas de milho (*Zea mays*), arroz (*Oryza sativa*), sorgo (*Sorghum bicolor*) e cana-de-açúcar (*Saccharum officinarum*) e plantas tropicais como bananeira e abacaxizeiro (BALDANI et al., 1986; BODDEY et al., 1995; MAGALHAES CRUZ et al., 2001; PIMENTEL et al., 1991). Diversos estudos têm mostrado que este potencial concentra-se na capacidade deste microorganismo sintetizar o fitormônio ácido-3-indol acético (AIA) que pode

afetar a morfologia e o desenvolvimento da planta colonizada, e também por sua capacidade de reduzir o nitrogênio gasoso à amônia que poderia ser disponibilizada para as plantas como fonte de nitrogênio. (BALDANI; BALDANI, 2005). O genoma de *H. seropedicae* estirpe SmR1 foi completamente sequenciado e anotado pelo consórcio GENOPAR¹ e publicado no ano de 2011 (PEDROSA et al., 2011). Foram identificadas 4.735 *open reading frames* (ORFs) em seu único cromossomo circular que contém 5.513.887 pares de nucleotídeos. Foram identificados 3 operons de rRNA completos e sequências codificadoras para 55 tRNAs (PEDROSA et al., 2011). Durante esse trabalho, o processo de anotação de genes não identificou a presença de nenhuma molécula de ncRNA.

1 <http://www.genopar.ufpr.br>

4. MATERIAIS E MÉTODOS

4.1 Conjunto de dados

4.1.1 Genoma

Os dados de sequência e anotação do genoma da bactéria *Herbaspirillum seropedicae* estirpe SmR1 foram obtidos através do banco de dados do *National Center for Biotechnology Information Search* (NCBI) sob a referência NC_014323.1.

4.1.2 Sequências de ncRNA

As sequências de ncRNAs *cis* e *trans* foram obtidas da base de dados *online* RFAM² que contém uma coleção de famílias de RNA, como por exemplo, *trans* e *cis* ncRNA (sRNA e *Riboswitch*), introns e RNAs leader. As sequências de *Riboswitches* foram obtidas do banco de dados RIBEX³.

4.1.3 Dados de RNA-seq

As leituras derivadas dos experimentos de RNA-seq foram gentilmente cedidas pelo Núcleo de Fixação Biológica de Nitrogênio. Os experimentos referentes ao cultivo da bactéria nas condições controle e na presença do flavonóide naringenina, extração de RNA e sequenciamento de cDNA na plataforma SOLiD foram realizados pela Dra. Michelle Zibetti Tadra Sfeir (TADRA-SFEIR, 2011). Os experimentos de transcriptoma da bactéria *H. seropedicae* SmR1 associada à planta foram realizados pelo aluno de doutorado Eduardo Balsanelli.

² <http://rfam.sanger.ac.uk/>

³ <http://132.248.32.45/cgi-bin/ribex.cgi>

4.2 Sistema operacional

Os trabalhos foram realizados utilizando a plataforma GNU/Linux com a distribuição CentOS 5.7⁴. Esse sistema é uma distribuição binariamente compatível com o Red Hat Enterprise⁵. Todos os softwares estão disponíveis livremente na WEB.

4.3 Linguagens de Programação

Para a geração, formatação e análise dos dados, foram desenvolvidos scripts auxiliares escritos em linguagem Bash, um interpretador de comandos largamente utilizado em plataforma Unix e GNU/Linux. A linguagem Bash⁶ foi escolhida devido à facilidade em utilizar expressões regulares, o que torna a manipulação de dados genômicos facilitada.

4.4 Ferramentas de Bioinformática

As seguintes ferramentas de bioinformática foram utilizadas:

4.4.1 Artemis

Programa para edição e anotação dos ncRNAs candidatos. Com a ferramenta BamView (CARVER et al., 2010), integrada ao software Artemis, é possível visualizar os alinhamentos dos reads de RNA-seq em um arquivo no formato BAM. Adicionalmente, valores de RPKM podem ser calculados para cada ncRNA candidato.

⁴ <http://www.centos.org/>

⁵ <http://www.br.redhat.com/products/rhel/>

⁶ <http://www.gnu.org/software/bash/>

4.4.2 CuffLinks

Identifica regiões expressas no genoma baseando-se em alinhamento de leituras derivados de RNA-seq. Torna possível a identificação de ncRNAs que não possuem características genômicas identificadas, como promotor e terminador, através da análise das regiões que estão sendo transcritas e que não correspondem à ORFs (TRAPNELL et al., 2010).

4.4.3 cmsearch

Procura de similaridade em estrutura primária e secundária dos ncRNA em uma base de dados. A base de dados utilizada foi obtida do site RFAM em 21/11/2011. Um *script* escrito na linguagem bash foi utilizado para a automação do processo (apêndice 3).

4.4.4 GSAlgorithm

Realiza busca de ncRNAs em regiões intergênicas através da identificação de características genômicas como presença de promotor fator sigma70 e terminador Rho-Independente (VALVERDE et al., 2008). Tais características são localizadas com o auxílio das ferramentas externas: NNPP (REESE, 2001) para predição de promotores e TranstermHP Server (KINGSFORD; AYANBULE; SALZBERG, 2007) para predição de terminadores de transcrição.

4.4.4.1 Predição de Promotores

Para a predição de promotores fator sigma 70, o programa GSAlgorithm utiliza a ferramenta externa NNPP (VALVERDE et al., 2008). Baseado em um servidor web, essa ferramenta utiliza rede neural para a predição de promotor. A ferramenta NNPP foi executada utilizando todos os parâmetros padrões com um *score* mínimo de 0.8.

4.4.4.2 Predição de Terminadores

Para a predição de terminadores Rho-independentes foi utilizada a ferramenta TarstermHP. Esse método computacional procura em genomas de procariotos grampos com baixa energia livre e seguidos de uma sequência de nucleotídeos de timina conforme está representado na FIGURA 5.

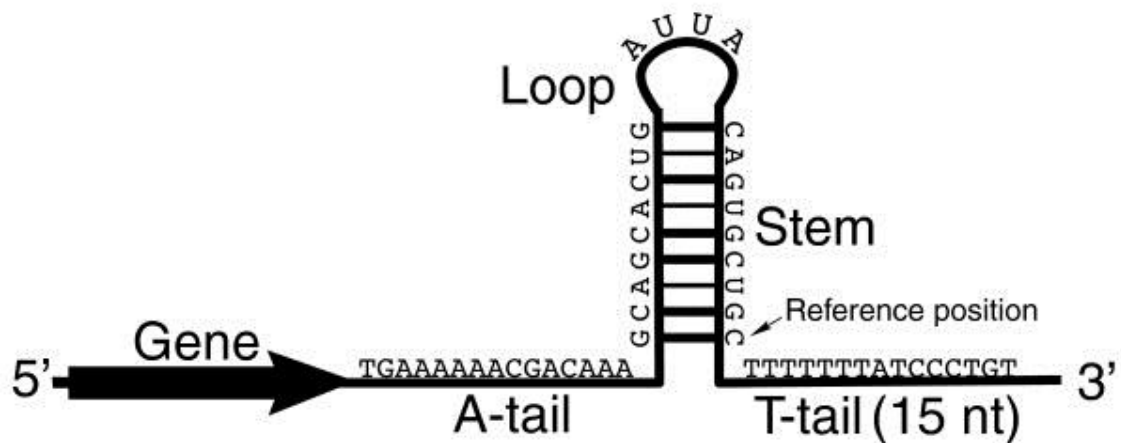


FIGURA 5: Esquema de um grampo terminador Rho-Independente. O terminador consiste de um grampo (*stem-loop*) seguido de uma sequência rica em nucleotídeos de timina (T-tail).Fonte: (KINGSFORD et al., 2007)

4.4.5 (non-coding RNA characterization) nocoRNAC

Realiza busca de ncRNAs em qualquer região de genomas bacterianos, incluindo regiões antisenso (HERBIG; NIESELT, 2011). Para tanto, prediz a presença de promotor detectando regiões de desestabilização no DNA e a presença de terminador Rho-Independente através da ferramenta TranstermHP (KINGSFORD et al., 2007).

A seguinte linha de comando foi executada:

```
Java -X mx10G -jar nocoRNAC.jar -genomeFastaFile genoma.fasta -proteintableFile CP002039.ptt -siddSites -importTransTerm outTransterm.out -gffOutFile analises.out analises_NocoRNAC.out -overwrite
```

4.4.5.1 Predição de Promotores

Para a predição de promotores, uma abordagem denominada de SIDD (*Stress Induced Duplex Destabilization*) foi utilizada. Nele, a fita de DNA é checada para verificar a existências de regiões possíveis de serem desestabilizadas. Esse método se baseia no conceito de que uma sequência de promotor é desestabilizada durante o processo de transcrição. A ferramenta nocoRNAC é capaz de localizar padrões de desestabilização semelhantes aos que ocorrem em promotores caracterizados (BENHAM; BI, 2004).

4.4.5.2 Predição de Terminadores

A ferramenta TranstermHP foi utilizada para a predição de terminadores Rho-Independentes. Para maiores detalhes ver item 4.4.4.1.

4.4.6 gsmapper

Programa para o mapeamento de *reads* de experimentos de RNA-seq contra um genoma de referência (RUMBLE et al., 2009). O programa está contido no pacote de ferramentas Shrimp.

4.4.7 SamTools

É um pacote de utilitários para a manipulação de mapeamentos em formato SAM (LI et al., 2009). Neste trabalho, os seguintes comandos foram utilizados:

1) Conversão do mapeamento de SAM para BAM

samtools view -bS arquivo.SAM > arquivo.BAM

2) Alinhamento das coordenadas

samtools sort arquivo.BAM arquivobam.sorted

3) Classifica o índice para acelerar o acesso ao arquivo

samtools index arquivo.sorted.bam

4.5 Hardware

Todas as análises foram realizadas utilizando um computador servidor de alto desempenho com as seguintes configurações:

Kernel – 2.6.18

Arquitetura – 64 bits

Processadores – Intel i7 950 3.07 GHz

Memória – 24GB

4.6 Transcriptoma

Todas as etapas realizadas em laboratório para as condições controle e na presença do flavonóide naringenina, como cultivo da bactéria, extração de RNA e sequenciamento de cDNA, foram desenvolvidas pela Dra. Michelle Zibetti Tadra Sfeir. Os experimentos de transcriptoma da bactéria associada à planta foram realizados pelo doutorando Eduardo Balsanelli. Para a verificação de detalhes técnicos, consultar (TADRA-SFEIR, 2011). Foram realizadas extrações totais de RNA de bactérias estirpe SmR1 de *H. seropedicae* nas seguintes condições de cultivo:

1) Controle (condição em que a bactéria não fixa nitrogênio atmosférico):

Herbaspirillum seropedicae SmR1 cultivado em meio NFbHPN contendo malato

como fonte de carbono. Amônio 20 mM. As células foram coletadas após 6 horas de cultivo – D.O._{600nm}= 0,8.

2) Presença do flavonóide Naringenina: *Herbaspirillum seropedicae* SmR1 cultivado em meio NFbHPN malato com 100 µmol/L de naringenina. Células coletadas após 6 horas de cultivo – D.O._{600nm}= 0,8.

3) Associada a plantas de milho híbrido Pioneer variedade 30F53: Células de *H. seropedicae* SmR1 extraídas de plântulas de milho após um e três dias da inoculação.

Para a extração de RNA foi utilizado o kit de isolamento de RNA Ribopure Bacteria (Ambion) de acordo com o protocolo sugerido pelo fabricante. O RNA total foi depletado de RNA ribossomal utilizando o kit MICROBExpress™ (Ambion) de acordo com as instruções do fabricante. Para a construção das bibliotecas e sequenciamento de cDNA foi utilizado o kit SOLiD™ Whole Transcriptome Analysis Kit seguindo as indicações do fabricante. Os fragmentos foram sequenciados em aparelho SOLiD 4 (Life Technology).

4.6.1 Mapeamento de Leituras de RNA-seq

Para o mapeamento das leituras contra o genoma alvo, as sequências de rRNA foram ocultadas utilizando a ferramenta *cross-match* (GORDON; ABAJIAN; GREEN, 1998). Dessa forma, o genoma de referência passou a ser o genoma de *H. seropedicae* SmR1 com todos os nucleotídeos das sequências de rRNAs trocados pela letra N. Antes do alinhamento das leituras de RNA-seq os 15 nucleotídeos da região terminal da leitura foram aparadas (*trimming*) para 35 nucleotídeos utilizando um script escrito utilizando a linguagem Perl (Apêndice II). As leituras no formato csfasta obtidos do equipamento de sequenciamento foram mapeados utilizando a ferramenta *gmapper* contida no pacote SHRiMP (RUMBLE et al., 2009). O arquivo csfasta que é obtido contém apenas informações de *color space*. Dessa forma, ao invés de apenas uma base de nucleotídeo ser lida a cada ciclo, são medidas simultaneamente informações de duas bases de nucleotídeo.

O seguinte comando de execução foi utilizado para o mapeamento:

```
gmapper-cs arquivo.csfasta referencia.fasta -N8 -o1 -h80% -E 2>arquivo.log>arquivo.sam
```

Onde o parâmetro $-N$ é referente ao número de núcleos de processamento que foi utilizado pelo computador, $-o$ seta o número de vezes que um *read* pode ser alinhado, $-h$ porcentagem de homologia entre as sequencias e $-E$ permite o arquivo de saída do alinhamento no formato *sam*. *Sam* é um formato de alinhamento compactado de *reads* curtos.

4.6.2 Análise do transcriptoma

Para a análise e normalização das amostras, inicialmente foram excluídas as leituras correspondentes a genes de tRNA. Após essa etapa, a normalização foi realizada pelo método de RPKM (*reads per kilobase of transcript per million mapped reads*) (MORTAZAVI et al., 2008), que estima o valor de expressão de um gene. Este método considera o tamanho da biblioteca, o tamanho dos genes e o número de leituras por gene, utilizando a fórmula:

$$R = \frac{10^9 C}{NL}$$

C = número de leituras por gene

L= tamanho do gene

N = tamanho da biblioteca (número total de leituras por replicata biológica)

Para determinar o valor de cobertura, a seguinte expressão matemática foi utilizada:

$$X = \frac{C * 35}{L}$$

C = número de leituras por gene

L = tamanho do gene

35 = tamanho da leitura de RNA-seq após o processo de *trimming*

5. RESULTADOS

5.1 Identificação de ncRNAs utilizando os programas GSalgorithm e nocoRNAC

A identificação de ncRNAs no genoma de *H. seropedicae* SmR1 envolveu duas buscas independentes que se basearam em características comuns a maioria dos ncRNAs já descritos. Essa etapa foi realizada utilizando as seguintes ferramentas de bioinformática: GSalgorithm (VALVERDE et al., 2008) e nocoRNAC (HERBIG; NIESELT, 2011).

A ferramenta GSalgorithm procura por genes de ncRNAs presentes em regiões intergênicas de genomas bacterianos. O método computacional utilizado pelo programa, permite, além da identificação dos ncRNAs, o ordenamento dos candidatos encontrados de acordo com a confiabilidade da predição (VALVERDE et al., 2008). Para a realização destas tarefas, o programa GSalgorithm cruza os dados gerados por dois *softwares* preditores de promotor fator sigma 70, o NNPP e o FuzzNuc, e um *software* preditor de terminador rho-independente, o TranstermHP (KINGSFORD et al., 2007; REESE, 2001; RICE; LONGDEN; BLEASBY, 2000).

O programa nocoRNAC realiza uma busca por todo genoma de uma bactéria para verificar a existência de genes de ncRNAs, não estando, portanto, restrito a regiões intergênicas (HERBIG; NIESELT, 2011). Esta abordagem computacional utiliza as informações geradas pela ferramenta TranstermHP para a predição de possíveis terminadores rho-independente, e o modelo SIDD para o reconhecimento de sequências de promotores (BENHAM; BI, 2004; KINGSFORD et al., 2007). Os dados gerados são unidos e comparados gerando um lista de candidatos a genes de ncRNAs.

A aplicação dessas ferramentas resultou em uma grande população de sequências candidatas a genes de ncRNAs que estão listadas na TABELA 2. Todos esses candidatos possuem promotores com fator sigma e terminadores Rho-independentes identificados *in-silico*.

TABELA 2 - NCRNAS DETECTADOS NO GENOMA DE *H. SEROPEDICAE* SMR1 UTILIZANDO AS FERRAMENTAS GSALGORITHM E NOCORNAC.

FERRAMENTA	CANDIDATOS DE ncRNAS
GSalgorithm	180
nocoRNAC	769
Total	949

Dados de sequências obtidas de experimentos de sequenciamento do transcriptoma, ou RNA-seq, disponíveis no laboratório foram utilizados para validar os ncRNAs candidatos.

A extração total de RNA para a realização de experimentos de RNA-seq foi realizada com a bactéria *H. seropedicae* SmR1 cultivadas nas seguintes condições: meio NFbHPN (controle), meio NFbHPN + naringenina e meio NFbHPN e a bactéria associada à planta (item 4.6). O material, após as etapas experimentais (item 4.6), foi sequenciado no aparelho SOLiD 3 (Life Technology) (4.6).

As sequências obtidas dos experimentos de RNA-seq foram tratadas e mapeadas no genoma da bactéria *H. seropedicae* SmR1 utilizando o pacote de ferramentas SHRiMP (item 4.4.6). O arquivo de mapeamento foi modificado para o modo binário (extensão .bam) para um melhor desempenho nas análises (item 4.4.7). O número de leituras mapeadas provenientes da condição em que a bactéria estava associada à planta apresentou-se reduzido devido a remoção das leituras de rRNA manualmente na tentativa de obter um resultado com menos interferência. O número de sequências mapeadas em cada experimento de RNA-seq está apresentado na TABELA 3.

TABELA 3 – SEQUÊNCIAS OBTIDAS DAS BIBLIOTECAS GENÔMICAS

AMOSTRA	NÚMERO TOTAIS DE LEITURAS	NÚMERO DE LEITURAS MAPEADAS
Controle	35.470,533	26.723,519*
Presença de Naringenina	15.084,728	11.231,142*
Associada à Planta	22,091,857	1,145,546

* Presença de rRNA

De posse do mapeamento, verificou-se o nível de transcrição de cada candidato à ncRNA. Para quantificar o nível de expressão de um gene através de dados de RNA-seq, uma operação matemática denominada de RPKM normaliza o número de leituras que alinharam em um transcrito, pelo tamanho do gene e pelo total de leituras mapeadas na amostra (MORTAZAVI et al., 2008). Adicionalmente, foi calculado qual a cobertura que cada gene candidato à ncRNA possuía. Para isso, a expressão matemática de cobertura (item 4.6.3) foi utilizada. Os candidatos que obtiveram o valor de cobertura igual ou superior a cinco, em pelo menos uma das condições testadas, foram classificados como ncRNAs confirmados experimentalmente. Aqueles candidatos com valor de cobertura inferior a cinco foram considerados como possíveis genes de ncRNA sem expressão identificada. O número de ncRNA experimentalmente confirmados está mostrado na Tabela 4.

TABELA 4. NCRNAS PREDITOS COM OS PROGRAMAS GSALGORITHM E NOCORNAC E CONFIRMADOS UTILIZANDO DADOS DE RNA-SEQ.

MÉTODO DE PREDIÇÃO	NÚMERO DE PREDITOS	NCRNA CONFIRMADOS
GSalgorithm	180	29
nocoRNAC	769	48
Total*	949	77

*O valor total não considera *overlap* de resultados.

Todos os 77 ncRNAs, confirmados experimentalmente, foram anotados em um arquivo no formato gbk, junto com todos os genes codificadores de proteínas de *H. seropedicae* SmR1. Esse trabalho foi realizado com o auxílio da ferramenta de

bioinformática Artemis ⁷. A FIGURA 6 exemplifica visualização gráfica do arquivo gbk e do arquivo de mapeamento, na condição controle, carregados no *software* Artemis.

⁷ <http://www.sanger.ac.uk/resources/software/artemis/>

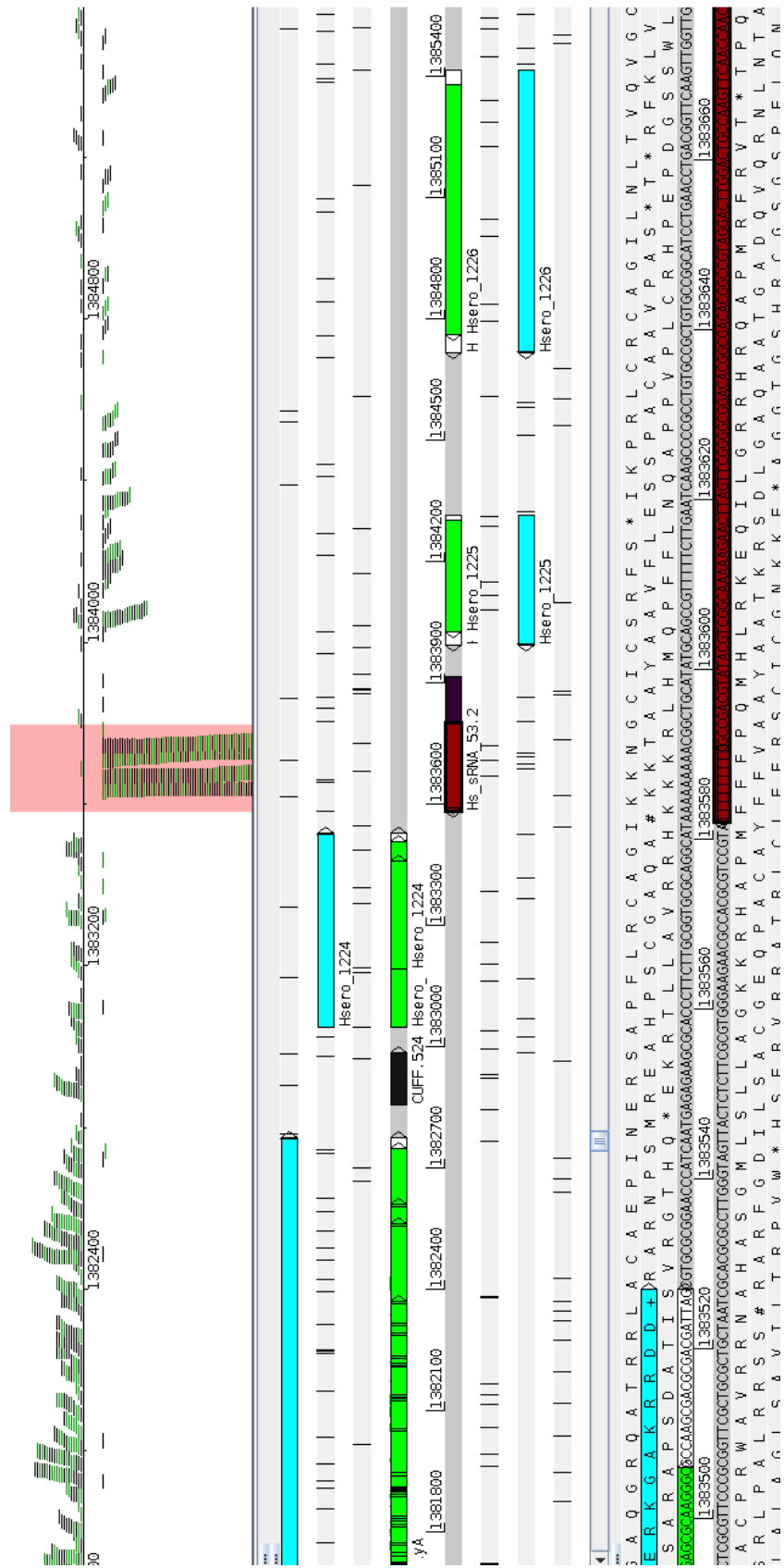


FIGURA 6 - Visualização do arquivo gbk e do arquivo de mapeamento no software Artemis. O gene Hs_sRNA53.2 é um ncRNA confirmado experimentalmente. A caixa em vermelho destaca as sequências derivadas do experimento de RNA-seq (condição controle), alinhadas ao ncRNA confirmando sua expressão.

5.2 Identificação de ncRNAs utilizando dados de RNA-seq

O algoritmo de montagem de transcriptoma *cufflinks* é adequado para a identificação de transcritos utilizando dados de RNA-seq (TRAPNELL et al., 2010). Como esta ferramenta não utiliza informações de anotações durante a análise, ela pode ser utilizada para o reconhecimento de qualquer espécie de transcrito, incluindo genes de ncRNAs. Por não necessitar de dados sobre sequências de promotores e de terminadores para a localização de transcritos, esta abordagem se mostra muito útil para a predição de novos genes de ncRNAs. Esta análise foi realizada somente com dados de RNA-seq da condição controle. O genoma usado para o mapeamento das leituras de RNA-seq foi tratado para que todas as ORFs e os 50 *pb* que as falqueiam fossem ocultadas. Foram realizados dois mapeamentos com leituras de RNA-seq da condição controle: o primeiro utilizando a fita de DNA principal e o segundo a fita de DNA reversa.

A aplicação dessa abordagem possibilitou a localização de 1395 regiões do genoma com significativos níveis de transcrição e que não correspondem a genes codificadores de proteínas. Essas regiões transcritas podem indicar a presença de genes de ncRNAs que estão sendo expressos na condição controle. A FIGURA 7 ilustra uma região do genoma de *H. seropedicae* SmR1 sendo transcrita que foi detectada pelo programa *cufflinks*.

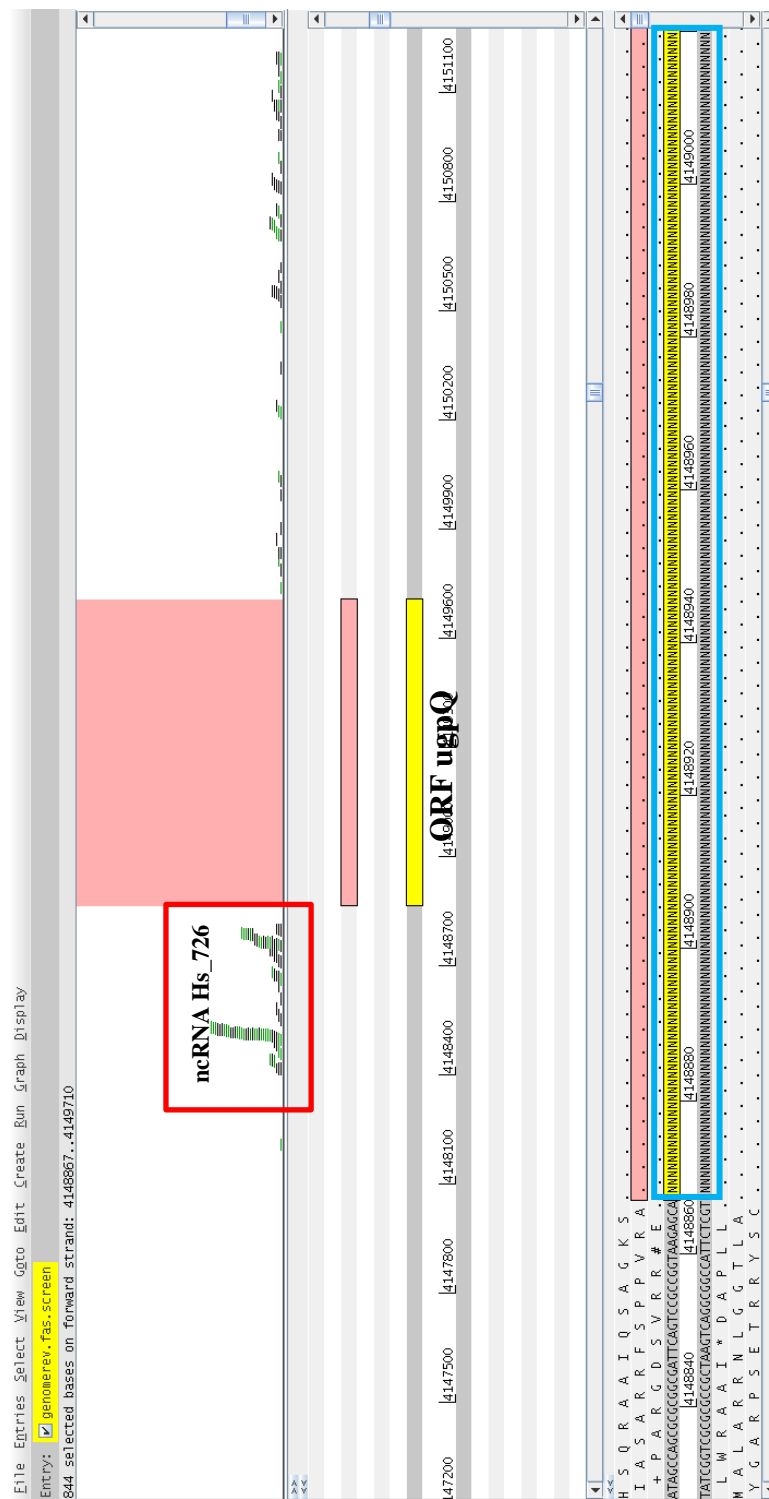


FIGURA 7 - Visualização do arquivo gbk e do arquivo de mapeamento no software Artemis. Destacado em vermelho está um ncRNA identificado pela ferramenta *cufflinks*. Downstream ao ncRNA está o gene *ugpQ*. O quadro em azul mostra que as ORFs e os nucleotídeos que as flanqueiam foram ocultadas corretamente na etapa anterior ao mapeamento.

O *cufflinks* quantifica o nível expressão dos candidatos através do cálculo de FPKM, uma medida equivalente ao valor de RPKM e que normaliza os dados pelo tamanho do gene e pelo total de leituras mapeadas na amostra (TRAPNELL et al., 2010). Como parâmetro de corte, o valor de cobertura de cada gene candidato foi calculado através da expressão matemática de cobertura. Os candidatos que obtiveram o valor de cobertura igual ou superior a cinco foram classificados como ncRNAs confirmados experimentalmente. Aqueles candidatos com valor de cobertura inferior a cinco, foram considerados como possíveis genes de ncRNA, porém sem expressão identificada na condição testada. A TABELA 6 contém o número de ncRNA experimentalmente confirmados.

TABELA 6. NCRNAS PREDITOS PELA FERRAMENTA CUFFLINKS E CONFIRMADOS EXPERIMENTALMENTE

MÉTODO DE PREDIÇÃO	NÚMERO DE PREDITOS	NCRNA CONFIRMADO
<i>Cufflinks</i>	1395	64

Todos os candidatos à ncRNA confirmados experimentalmente por essa abordagem foram incluídos no arquivo gbk criado anteriormente.

5.3 ncRNAs confirmados experimentalmente em *H. seropedicae* SmR1

A identificação de ncRNAs no genoma de *H. seropedicae* SmR1 envolveu três buscas independentes que foram realizadas utilizando ferramentas de bioinformática. A estratégia empregada está apresentada no fluxograma da FIGURA 8.

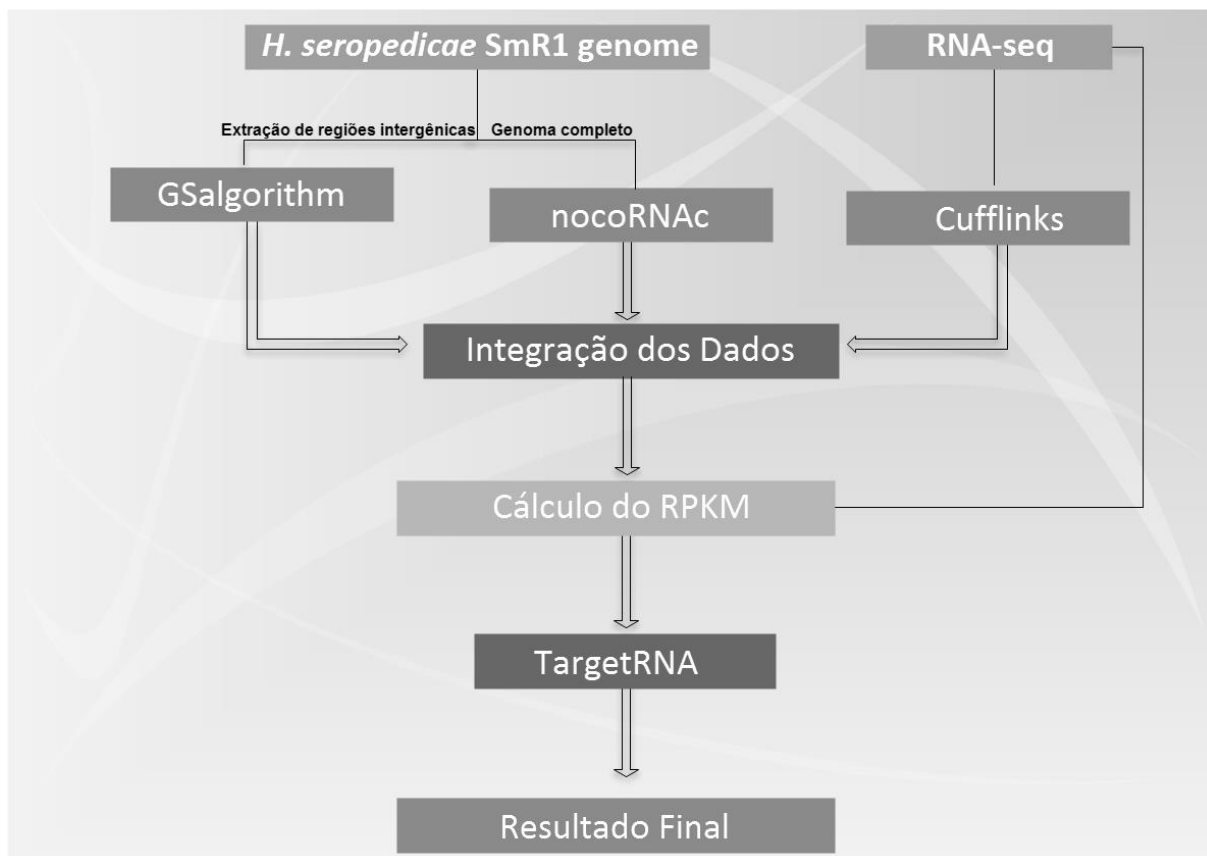


FIGURA 8 - Fluxo de trabalho utilizado para a identificação de ncRNAs na bactéria *Herbaspirillum seropedicae* estirpe SmR1.

As análises utilizando dados de RNA-seq, de três condições de crescimento diferentes, permitiram confirmar a presença de 98 ncRNAs. Os dados globais de validação estão apresentados na TABELA 7 e, neste trabalho, o maior número de ncRNAs validados experimentalmente foi observado na condição controle.

TABELA 7 – NÚMERO DE NCRNAS COM EXPRESSÃO CONFIRMADA EM TRÊS CONDIÇÕES EXPERIMENTAIS

AMOSTRA	NÚMERO DE NCRNAS CONFIRMADOS
Controle	92
Presença de Naringenina	75
Associada à Planta	37

Na FIGURA 9 está mostrada a comparação da quantidade de ncRNAs confirmados entre as diferentes condições experimentais. Dezesete ncRNAs apresentaram expressão somente em células cultivadas na condição controle, seis expressaram

somente quando a bactéria estava associada à planta, 44 expressaram tanto em células cultivadas na condição controle como em células cultivadas na presença de naringenina. Não foi observada expressão exclusiva quando as células foram cultivadas na presença de naringenina e 31 ncRNAs foram expressos nas três condições experimentais.

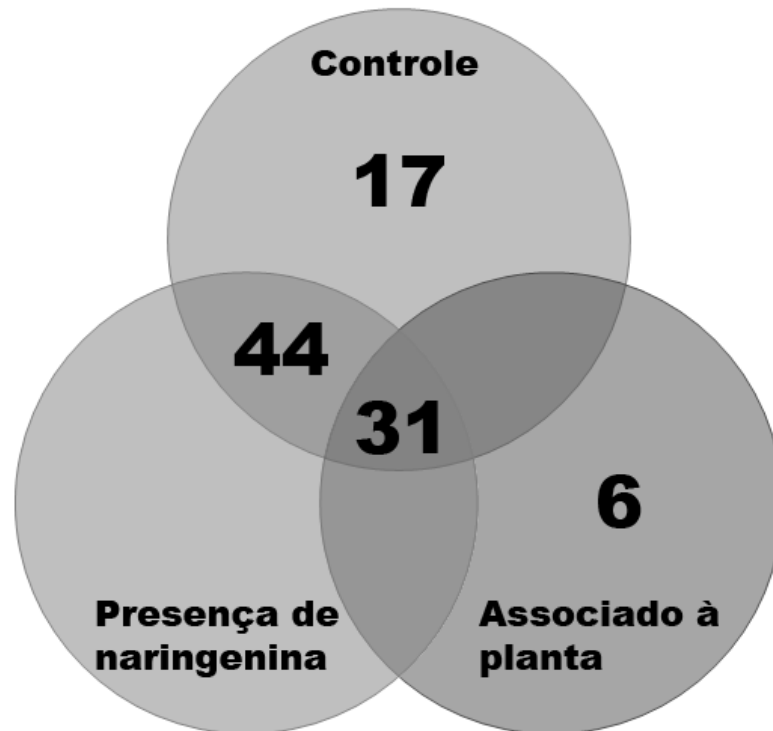


FIGURA 9: Diagrama de Venn representando o número de ncRNAs que possuem a expressão confirmada em cada uma das três condições de crescimento.

Dentre os 98 ncRNAs validados experimentalmente, 55 foram identificados na fita de DNA senso e 43 na fita de DNA antisense. O tamanho médio dos ncRNAs encontrados foi de 192 pares de bases, desconsiderando a sequência promotora daqueles preditos pelas ferramentas nocoRNAc e Gsalgorithm. O menor ncRNA encontrado possui 49 pares de base e o maior 552 pares de bases. Na tabela 8 estão apresentados os detalhes de cada ncRNA encontrado e validado experimentalmente.

TABELA 8 – CANDIDATOS ENCONTRADOS NO GENOMA DE *H. SEROPEDICAE* SMR1 E CONFIRMADOS EXPERIMENTALMENTE UTILIZANDO DADOS DE RNA-SEQ DE TRÊS CONDIÇÕES DE CRESCIMENTO DISTINTAS.

NOME	POSIÇÃO INICIAL	POSIÇÃO FINAL	TAMANHO DA SEQ.	ORIENTAÇÃO	FERRAMENTA DE PREDIÇÃO
"CUFF.55"	205713	205798	85	>	CuffLinks
Hs_noco_471	219112	219229	117	>	nocoRNAc
"CUFF.111"	359683	359916	233	>	CuffLinks
"CUFF.124"	412442	412676	234	>	CuffLinks
Hs_noco_490	417221	417409	188	>	nocoRNAc/GS/cufflinks
Hs_noco_494	457262	457311	49	>	nocoRNAc/cufflinks
TPP	462470	462706	236	>	Gs/cufflinks
"CUFF.225"	539844	539917	73	<	CuffLinks
"CUFF.228"	541704	541796	92	<	CuffLinks
Hs_noco_498	551015	551235	220	>	nocoRNAc/GS/cufflinks
Hs_sRNA_19.2	612014	612436	422	<	Gs/cufflinks
"CUFF.223"	664741	664801	60	>	CuffLinks
"CUFF.251"	741007	741139	132	>	CuffLinks
"CUFF.273"	792569	792670	101	>	CuffLinks
Hs_sRNA_Gs10	808782	808981	199	<	Gs/cufflinks
Hs_noco_514	808928	809191	263	>	nocoRNA/Gs/cufflinks
Hs_noco_525	932018	932087	69	<	nocoRNAc/cufflinks
Hs_noco_526	935267	935546	279	>	nocoRNA/Gs/cufflinks
Hs_noco_530	957449	957626	177	>	nocoRNAc
Hs_noco_534	1067480	1067548	68	<	nocoRNAc/cufflinks
"CUFF.395"	1155828	1155922	94	>	CuffLinks
"CUFF.404"	1189098	1189333	235	>	CuffLinks
"CUFF.504"	1336439	1336564	125	>	CuffLinks
"CUFF.522"	1381282	1381335	53	>	CuffLinks
"CUFF.524"	1382857	1382981	124	>	CuffLinks
Hs_noco_551	1383591	1383913	322	<	nocoRNAc/GS/cufflinks
Hs_noco_555	1433080	1433239	159	>	nocoRNA/Gs/cufflinks
Hs_noco_560	1495976	1496389	413	<	nocoRNAc
Hs_noco_564	1587943	1588410	467	>	nocoRNA/Gs/cufflinks
Hs_noco_565	1596987	1597275	288	<	nocoRNA/Gs/cufflinks
Hs_noco_568	1636668	1636831	163	>	nocoRNAc/GS/cufflinks
Hs_sRNA_Gs36	1788675	1788860	185	<	Gs/cufflinks
"CUFF.689"	1867577	1867679	102	>	CuffLinks
"CUFF.690"	1880603	1880826	223	>	CuffLinks
Hs_sRNA_Gs6	1915816	1915944	128	>	Gs
"CUFF.738"	2011544	2011601	57	>	CuffLinks
Hs_noco_593	2032405	2032540	135	>	nocoRNA/Gs/cufflinks
Hs_noco_602	2140731	2141104	373	<	nocoRNAc/GS/cufflinks
Hs_noco_604	2177189	2177339	150	>	nocoRNAc/GS
"CUFF.829"	2225999	2226140	141	>	CuffLinks
Hs_sRNA_Gs12	2233545	2233793	248	<	Gs/cufflinks
Hs_sRNA_4.2	2234033	2234214	181	>	Gs/cufflinks

"CUFF.852"	2287397	2287464	67	>	CuffLinks
"CUFF.934"	2519484	2519726	242	>	CuffLinks
"CUFF.938"	2526174	2526304	130	>	CuffLinks
"CUFF.967"	2578050	2578397	347	>	CuffLinks
"CUFF.1011"	2800595	2800705	110	>	CuffLinks
Hs_noco_637	2815035	2815179	144	<	nocoRNAc
Hs_noco_640	2939684	2939824	140	>	nocoRNA/Gs/cufflinks
Hs_noco_644	2972948	2973103	155	>	nocoRNAc/GS
"CUFF.1094"	3160202	3160357	155	>	CuffLinks
Hs_noco_651	3173522	3173679	157	<	nocoRNAc/cufflinks
"CUFF.1027"	3181155	3181240	85	<	CuffLinks
"CUFF.1117"	3204507	3204585	78	>	CuffLinks
"CUFF.1142"	3303508	3303560	52	>	CuffLinks
"CUFF.1109"	3382646	3382739	93	<	CuffLinks
"CUFF.1106"	3382844	3382966	122	<	CuffLinks
"CUFF.1223"	3512060	3512227	167	>	CuffLinks
"CUFF.1158"	3553031	3553120	89	<	CuffLinks
"CUFF.1190"	3614107	3614194	87	<	CuffLinks
"CUFF.1194"	3630870	3630996	126	<	CuffLinks
Hs_noco_682	3633077	3633378	301	<	nocoRNAc/cufflinks
Hs_noco_698	3779834	3779991	157	<	nocoRNAc/cufflinks
"CUFF.598"	3781409	3781584	175	>	CuffLinks
Hs_noco_700	3791718	3791799	81	<	nocoRNAc/cufflinks
Hs_noco_704	3879771	3880105	334	>	nocoRNAc
Hs_noco_709	3928375	3928556	181	>	nocoRNAc/cufflinks
Hs_noco_714	3998440	3998528	88	<	nocoRNAc/GS/cufflinks
Hs_noco_719	4074941	4075009	68	<	nocoRNAc
Hs_noco_722	4105880	4106014	134	<	nocoRNAc/cufflinks
Hs_sRNA_2.1	4108897	4109042	145	<	Gs/cufflinks
Hs_noco_723	4127435	4127552	117	<	nocoRNAc/cufflinks
Hs_noco_724	4127685	4127782	97	<	nocoRNA/Gs/cufflinks
Hs_noco_725	4136562	4136795	233	<	nocoRNAc/GS/cufflinks
Hs_noco_726	4148338	4148890	552	<	nocoRNAc/cufflinks
Hs_noco_732	4239106	4239615	509	>	nocoRNAc/cufflinks
Hs_noco_742	4396086	4396487	401	>	nocoRNAc/cufflinks
"CUFF.1510"	4402051	4402283	232	<	CuffLinks
Hs_noco_744	4417512	4417615	103	<	nocoRNAc/cufflinks
Hs_noco_746	4428208	4428599	391	<	nocoRNAc/cufflinks
Hs_noco_748	4467236	4467750	514	>	nocoRNAc/cufflinks
Hs_noco_749	4473739	4474078	339	<	nocoRNAc/cufflinks
Hs_noco_750	4474802	4474942	140	>	nocoRNAc/cufflinks
"CUFF.1548"	4499669	4499780	111	<	CuffLinks
Hs_noco_757	4564951	4565134	183	<	nocoRNAc/GS/cufflinks
"CUFF.1570"	4587035	4587149	114	<	CuffLinks
Hs_noco_761	4618621	4618928	307	>	nocoRNAc/cufflinks
CUFF.1692	4834369	4834486	117	<	CuffLinks
Hs_noco_780	4896565	4896712	147	>	nocoRNAc/GS/cufflinks
"CUFF.1713"	4909594	4909823	229	<	CuffLinks
"CUFF.1716"	4925909	4926013	104	>	CuffLinks
"CUFF.1720"	4929103	4929157	54	<	CuffLinks
"CUFF.1751"	5000211	5000493	282	<	CuffLinks
Hs_noco_786	5019358	5019702	344	>	nocoRNAc/GS/cufflinks

Hs_noco_788	5026868	5027073	205	<	nocoRNAc
Hs_noco_802	5239869	5240191	322	>	nocoRNAc/cufflinks
Hs_noco_811	5335529	5335812	283	>	nocoRNAc/GS/cufflinks
Hs_noco_820	5397591	5398115	524	>	nocoRNAc

De acordo com a organização proposta neste trabalho, 77 ncRNAs foram classificados como trans-sRNA, 17 como cis-sRNA e 4 como *Riboswitch*. A representação completa do desempenho das ferramentas utilizadas na predição de cada classe de ncRNA está ilustrada na FIGURA 10.

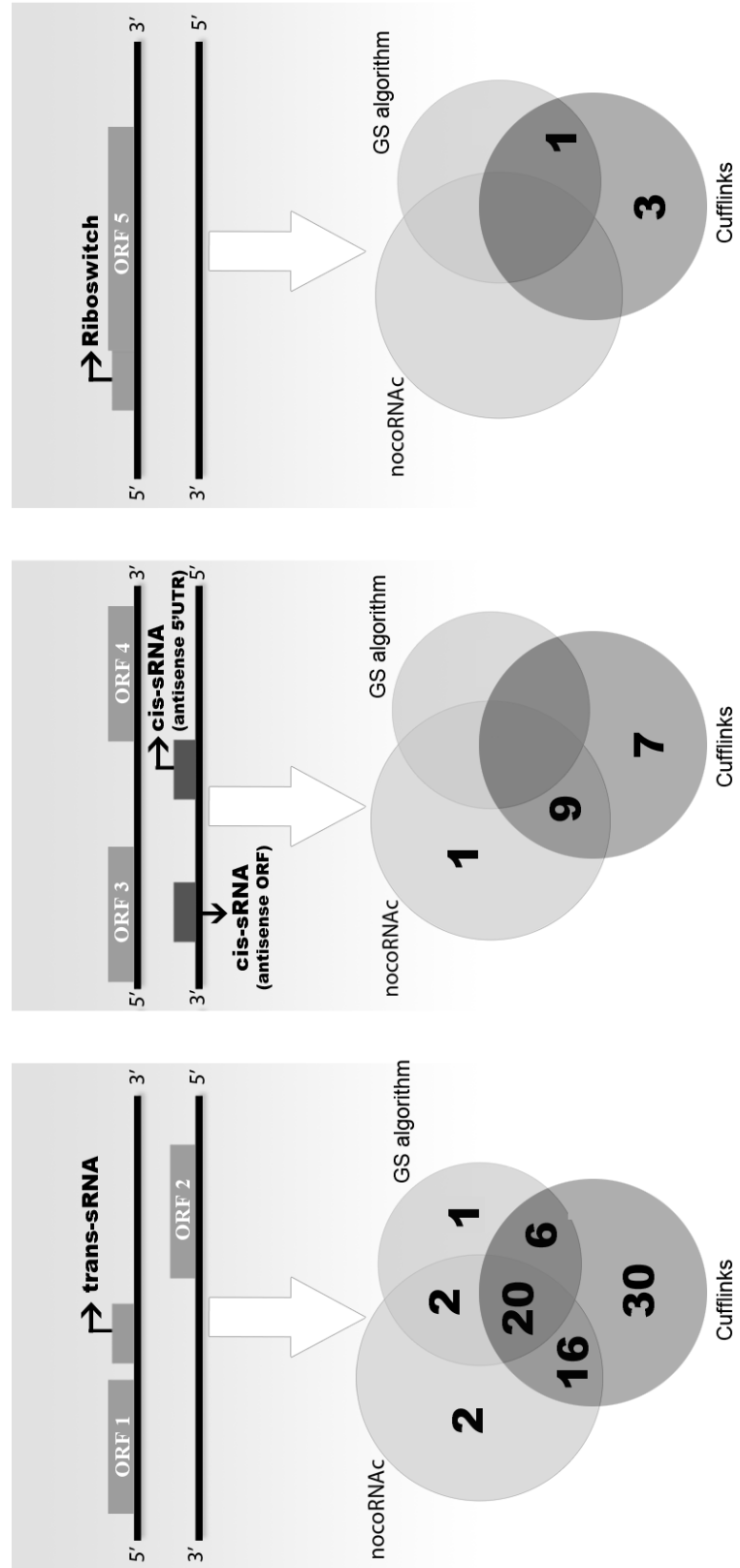


FIGURA 10 - Número de ncRNAs encontrados por 3 métodos de predição: nocoRNAC, GS algorithm e Cufflinks. Os ncRNAs estão organizados em trans-sRNA, cis-sRNA e *Riboswitch*.

5.4 Conservação dos ncRNAs

A busca por similaridade dos ncRNAs validados experimentalmente com outros já descritos e depositados em bancos de dados, produziu uma lista de oito moléculas, sendo elas:

(1) 6S/SsrS, que é expresso constitutivamente e inibe a transcrição através de sua interação direta com o fator sigma 70 da holoenzima RNA polimerase (WILLKOMM; HARTMANN, 2005).

(2) 4.5 S, presente em todos os procariotos, está envolvido em processos de secreção de proteínas (BROWN, 1991).

(3) tmRNA, com função dupla de tRNA e RNA mensageiro, desempenha um importante papel em processos como degradação de mRNA defeituosos (KEILER, 2008).

(4) Intron_gpl, uma ribozima que catalisa seu próprio *splicing* a partir de seu RNA precursor (NIELSEN; JOHANSEN, 2009).

(5) TPP *riboswitch*, associado com a biossíntese de tiamina (Vitamina B1) (MIRANDA-RIOS; NAVARRO; SOBERON, 2001).

(6) FMN *riboswitch*, que é necessário para o controle transcricional de genes na biossíntese de riboflavina (PETERS et al., 2009).

(7) Os elementos RLE0100 e RLE0038, não possuem funções conhecidas e descritas na literatura, porém estão relacionados com vias metabólicas presentes no KEGG e apresentam em sua vizinhança os mesmos genes em várias espécies bacterianas. Essas características fornecem evidências para que o banco de dados RIBEX classifique essas moléculas como possíveis *Riboswitches*.

Na TABELA 9 estão apresentadas as características dos ncRNAs conservados em outras espécies de bactérias.

TABELA 9 – NCRNAS CONSERVADOS EM OUTRAS ESPÉCIES DE BACTÉRIAS

NOME	CLASSE	TAMANHO	LOG2 RPKM	DATABASE
6S/SsrS	trans-sRNA	322 nt	16,5	RFAM
4.5 S	trans-sRNA	145 nt	13,8	RFAM
tmRNA	trans-sRNA	509 nt	18,2	RFAM
Intron_gpl	cis-sRNA	467 nt	11,2	RFAM
FMN	<i>Riboswitch</i>	141 nt	9,2	RIBEX
TPP Element	<i>Riboswitch</i>	292 nt	8,4	RIBEX
RLE0100	Putative Riboswitch-like Elements	114 nt	10,1	RIBEX
RLE0038	Putative Riboswitch-like Elements	110 nt	8,2	RIBEX

Os valores de RPKM determinados na condição controle e convertidos para log2 por conveniência

Através de uma busca por similaridade de sequência utilizando o programa blastn e o banco de dados nr/nt (*nucleotide collection*) do NCBI foi possível observar que as sequências dos ncRNAs validados experimentalmente estão conservadas em outras espécies de bactérias. Na FIGURA 11 pode ser observado o número de ncRNAs de *H. seropedicae* SmR1 que apresentam similaridade com sequências de outras espécies de bactérias. O valor mínimo de identidade tolerada foi de 70% no programa BlastN.

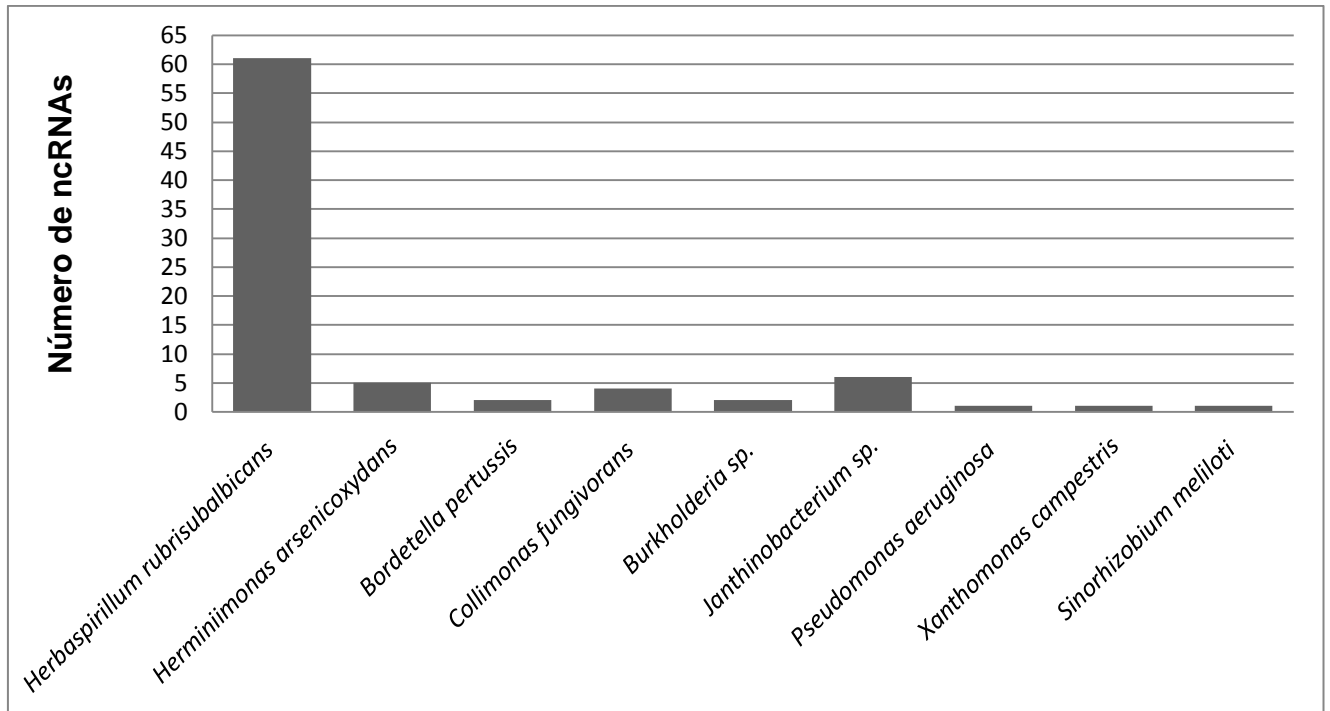


FIGURA 11 - Similaridade entre os ncRNAs de *H.seropedicae* SmR1 e seqüências de outras espécies bacterianas depositadas no banco de dados do NCBI.

5.5 Análise de dois prováveis *Riboswitches*: RLE0100 e RLE0038

Análises utilizando dados de RNA-seq de duas condições, controle e naringenina, mostram que os *Riboswitches* RLE0100 e RLE0038 podem estar regulando negativamente a expressão dos genes de mRNA localizados a jusante. Na FIGURA 12 está apresentada uma sugestão de mecanismo de regulação, no qual o *Riboswitch* estaria impedindo que o gene de mRNA seja transcrito completamente. Essa sugestão se baseia no número de leituras de RNA-seq que mapearam no *Riboswitch* e no respectivo gene de mRNA regulado. Neste, o número de leituras mapeadas é menor, indicando que a quantidade de transcrito produzido está reduzida.

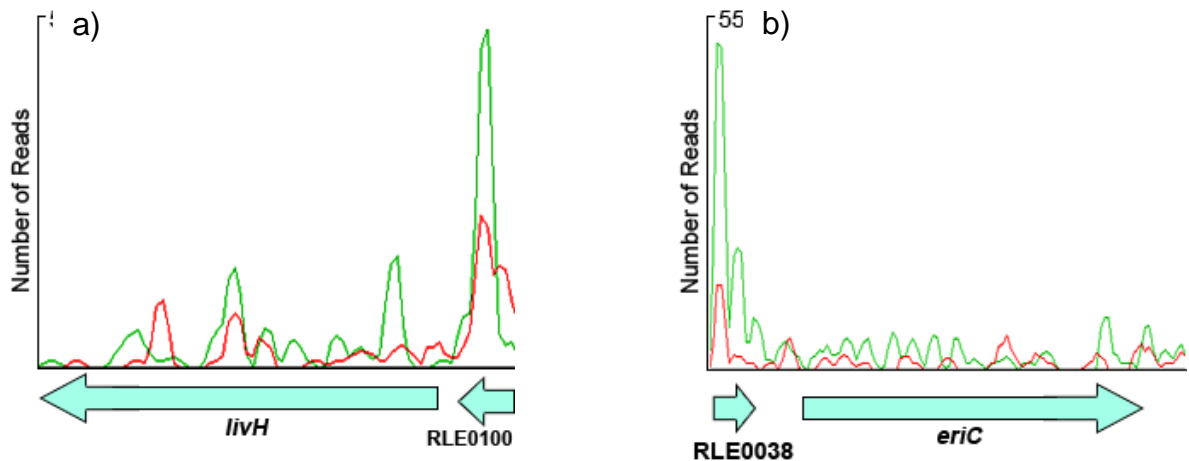


FIGURA 12 - Regulação de mRNAs por dois possíveis *Riboswitches*. As linhas em verde e vermelho representam as leituras mapeadas de RNA-seq nas condições controle e na presença de naringenina respectivamente. Os dados utilizados foram obtidos de experimentos independentes e por isso as escalas são diferentes

5.7 Alvos dos ncRNAs de *H. seropedicae* SmR1 confirmados experimentalmente

A rede de interações entre os ncRNAs e seus mRNAs alvos foi sugerida pelo programa TargetRNA. Nele, cada ncRNAs foi comparado, através de alinhamento local, com todos as sequências de mRNA de *Herbaspirillum seropedicae* SmR1. O programa organiza os possíveis alvos de acordo com uma pontuação interna denominada de *score*. Utilizando a ferramenta Cytoscape foi possível visualizar graficamente as interação entre todos os ncRNAs se seus alvos. Na Figura 13 é apresentada a complexa rede de interação entre os 98 ncRNAs e seus alvos. Não ficou evidenciado um padrão de interação entre as moléculas. Dessa forma, alguns ncRNAs se relacionam apenas com um alvo, enquanto outros possuem até oito alvos preditos, o número máximo que pode ser identificado pela ferramenta TargetRNA nesta análise.

A FIGURA 14 apresenta a interação de dois ncRNAs e seus alvos de forma detalhada. Os ncRNAs *ncRNA_Hs230* e *CUFF1716* compartilham um alvo em comum, o mRNA *Hsero3656*. Esse mRNA codifica para uma proteína com função de Lipase. Observa-se ainda, que o *CUFF1716* regula outros 3 alvos, dentre eles o mRNA *recX* que codifica uma proteína envolvida na resposta SOS (GALVAO et al.,

2003; GALVAO et al., 2012). A rede de interação do *ncRNA_Hs230* por sua vez, mostra que outros nove mRNA podem estar sendo regulados por ele.

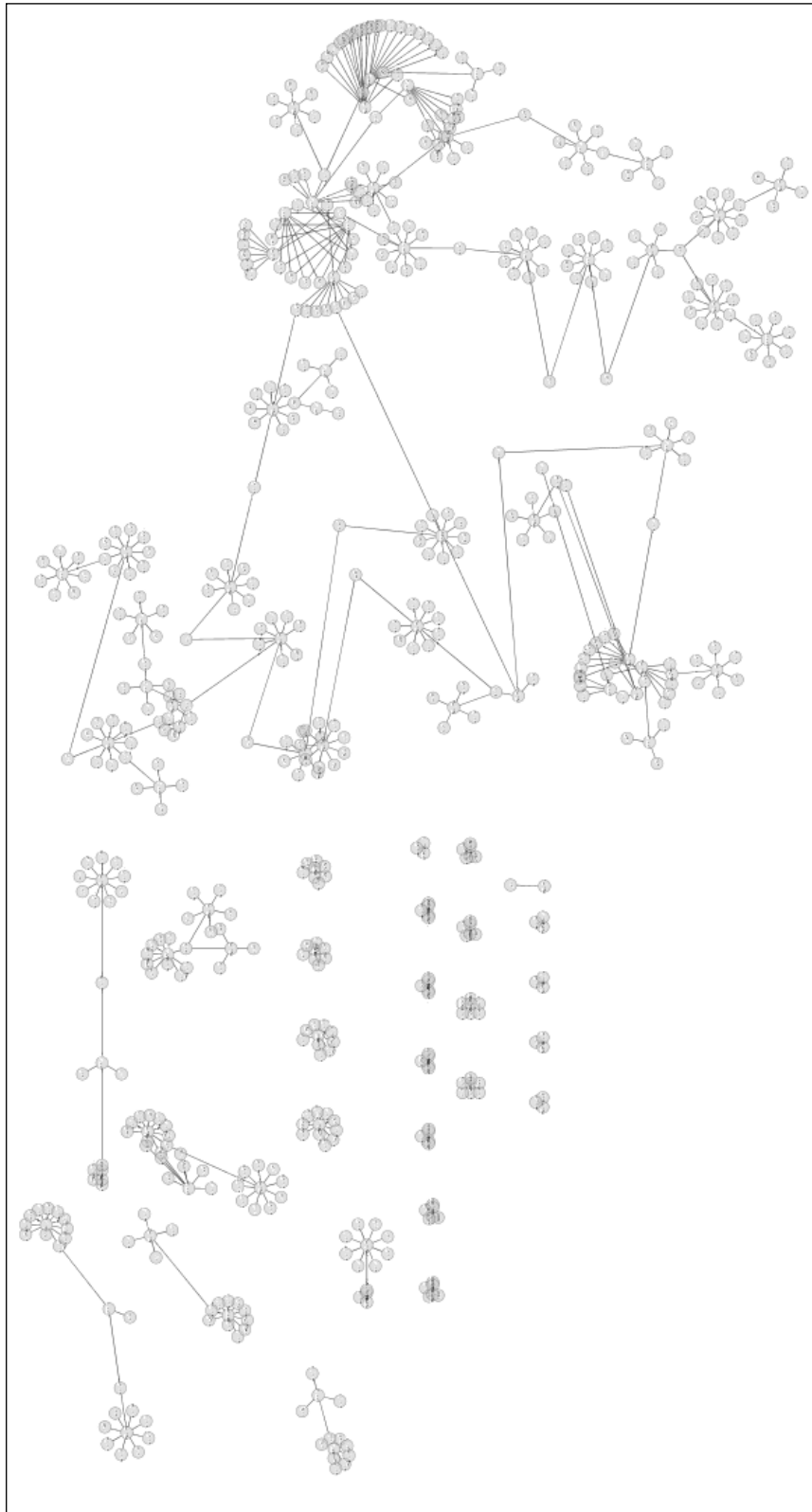


FIGURA 13 - Rede de interação entre os ncRNAs e seus mRNAs alvos.

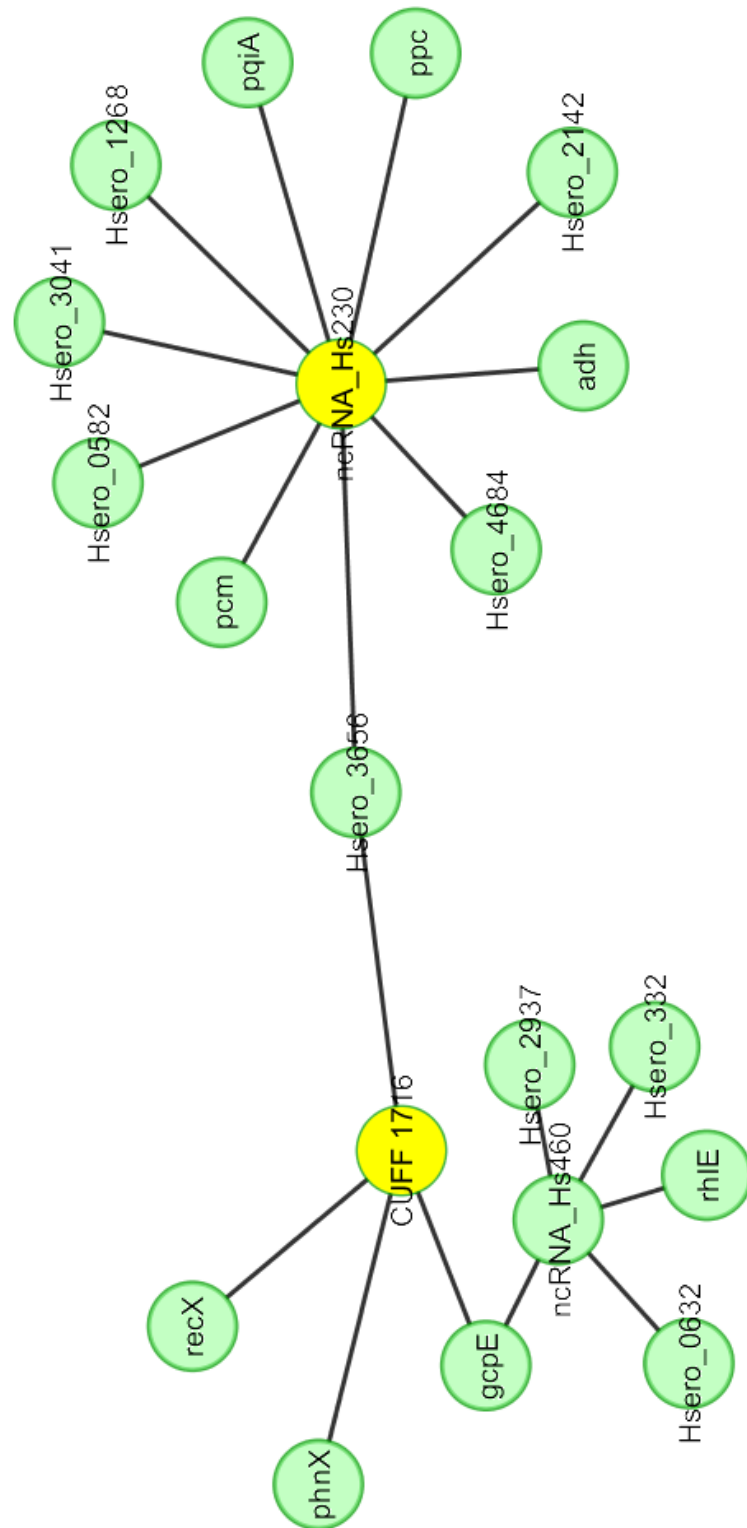


FIGURA 14: Rede de interação entre dois ncRNAs e seus mRNAs alvos. Destacado em amarelo estão os ncRNAs identificados na condição de cultivo controle e, em verde, os possíveis mRNAs alvos.

6 DISCUSSÃO

Os ncRNAs são moléculas capazes de modular a expressão de um gene em nível de transcrição ou de tradução. Respondendo a alterações ambientais, regulam seus alvos de mRNA através de pareamento de bases, ou interagem diretamente com proteínas regulando-as (REPOILA; DARFEUILLE, 2009). De acordo com sua localização no genoma podem ser classificadas em trans-sRNA, cis-sRNA e *Riboswitch* (WATERS; STORZ, 2009). A classe das Beta-proteobactérias apresenta poucos estudos no que se refere à identificação de ncRNAs (COENYE et al., 2007). Em *H. seropedicae* estirpe SmR1 nenhum artigo completo confirmando a existência dessas moléculas foi publicado até o momento, porém a existência destas moléculas foi já sugerida através de abordagem bioinformática (PEDROSA et al., 2011; WILLEMANN, 2010).

Com o advento de métodos computacionais eficientes e utilizados na predição *in silico* de ncRNA, combinados com métodos experimentais como RNA-seq, o número de ncRNAs identificados em bactérias vem crescendo nos últimos anos (STORZ; ALTUVIA; WASSARMAN, 2005).

Dessa forma, considerando a inexistência de evidências experimentais que comprovem a expressão de ncRNAs em *H. seropedicae* SmR1, e sabendo do envolvimento dessas moléculas na regulação de importantes vias metabólicas, procuramos investigar através de métodos computacionais e experimentais a presença de ncRNAs no genoma dessa bactéria. Também procuramos entender quais genes de mRNA estão, possivelmente, sendo modulados por ncRNAs. Para isso, relacionamos os resultados obtidos com os bancos de dados especializados em armazenar informações sobre ncRNAs caracterizados em outras espécies de bactérias.

Utilizando três abordagens computacionais distintas, foram encontradas 2344 moléculas de ncRNA no genoma de *H. seropedicae* SmR1. Após a confirmação da transcrição desses candidatos através de dados de RNA-seq obtidos a partir de três condições experimentais, um conjunto de 98 ncRNAs foi selecionado. Este conjunto está composto por genes que apresentaram níveis de transcrição igual ou superior ao valor mínimo pré-estabelecido através de experimentos de bioinformática com dados de RNA-seq.

Entre as abordagens utilizadas para as predições de ncRNAs, a aplicação da ferramenta *Cufflinks* resultou no maior número dessas moléculas, sendo 1395 preditos e 64 confirmados experimentalmente. Essa ferramenta utiliza o mapeamento de leituras de RNA-seq para localizar regiões do genoma sendo transcritas, porém livres de ORFs (TRAPNELL et al., 2010). Devido ao fato de não necessitar de informações sobre a sequência de promotor fator sigma 70 e de terminador Rho-independente para realizar as predições, esse programa torna possível a identificação de ncRNAs com promotores de fator sigma diferentes, como os promotores sigma 54 e sigma S (FILIATRAULT et al., 2010), ou ainda ncRNAs com diferentes mecanismos de terminação de transcrição, como terminadores Rho-dependentes (GARDNER; BARQUIST; et al., 2011; OTT et al., 2012).

A utilização da ferramenta de predição Gsalgorithm permitiu identificar um menor número de ncRNAs, sendo 26 confirmados experimentalmente de um grupo de 180 preditos. Por restringir a busca a regiões intergênicas, essa abordagem não localiza moléculas na posição antisense a ORFs (VALVERDE et al., 2008). Nossos resultados mostraram que apesar do menor número de predições de ncRNAs (Figura 10), o Gsalgorithm pode ser usado como mais um indício de que as predições realizadas por outras ferramentas estejam corretas. Além disso, fornece informações sobre a sequência do promotor, quando sigma 70, e a sequência de terminador Rho-independente.

O programa nocoRNAC combina a posição de possíveis promotores com a posição de terminadores Rho-independentes para identificar ncRNAs em genomas bacterianos (HERBIG; NIESELT, 2011). Com a capacidade de analisar o genoma de forma completa, não está restrito a regiões intergênicas. Nossa análise utilizando essa ferramenta resultou em 769 ncRNAs encontrados, os quais 44 tiveram a expressão comprovada. O modelo para a predição de promotores empregado por esse programa se baseia na localização de regiões do DNA possíveis de serem desestabilizadas, e não está restrito a localização de promotores com o fator sigma 70. Isso se torna uma vantagem, pois já foi reportado que alguns ncRNAs apresentam promotores com fator sigma S e sigma 54 (FILIATRAULT et al., 2010).

Por representar o repertório completo de transcritos, o sequenciamento do transcriptoma é uma abordagem experimental adequada para confirmar e quantificar a presença de ncRNAs em genomas bacterianos (KOGENARU et al., 2012). Neste trabalho, utilizando o RNA-seq de três condições de cultivo, foi possível verificar que

essa metodologia - predição *in silico* e confirmação experimental com transcriptoma - facilita a identificação de ncRNA. É bem estabelecido na literatura que muitos ncRNAs são transcritos para a adaptação da bactéria às alterações no ambiente (REPOILA; DARFEUILLE, 2009; STORZ et al., 2005; VERCRUYSSSE et al., 2010). Diante disso, muitos ncRNAs por nós preditos, que não tiveram a expressão identificada nas condições testadas, sejam transcritos em outras condições de cultivo específicas.

A predição dos possíveis mRNAs alvos para os 98 ncRNAs confirmados experimentalmente gerou uma extensa lista que sugere o compartilhamento de alvos por ncRNAs. É interessante notar que devido à interação entre o ncRNA e o mRNA envolver poucos pares de bases, e esse pareamento comumente ocorrer de forma imperfeita, cada ncRNA pode regular múltiplos alvos o que torna a análise complexa e passível de erros. O emprego de técnicas de predição de alvos *in silico* proporciona um indício da interação entre um ncRNA e seu alvo, porém experimentos laboratoriais são necessários para a exata elucidação dos processos regulatórios.

Em nossas análises, foi constatado que 8 ncRNAs possuem similaridade com ncRNAs depositados em bancos de dados. Embora a literatura sugira que a sequência primária dos ncRNAs seja pouco conservada entre espécies e a estrutura secundária de difícil modelagem para posterior comparação (BARQUIST, 2012), foi possível montar uma tabela (Tabela 9) mostrando ncRNAs presentes em *H. seropedicae* SmR1 e conservado em outras espécies de bactérias. De fato, os ncRNAs conservados e identificados são aqueles classificados como ncRNAs *housekeeping* por serem expressos constitutivamente e essenciais para a sobrevivência do microrganismo (LIOLIOU et al., 2012). Dessa forma, devido à sua importância, é esperado que esses ncRNAs sejam conservados durante a história evolutiva das bactérias.

As leituras de RNA-seq das bactérias associadas à planta confirmaram experimentalmente 6 ncRNAs exclusivos dessa condição. A predição de alvos realizada para esses ncRNAs não forneceu informações suficientes para o entendimento de quais vias metabólicas essas moléculas estão envolvidas. Porém, devido à condição de cultivo, é provável que a bactéria quando coletada estivesse fixando nitrogênio, numa condição diferente da condição controle, o que pode indicar um possível envolvimento desses ncRNAs com o processo de fixação biológica de

nitrogênio ou, ainda, que estes sejam ncRNAs regulatórios que são expressos durante o processo de interação da bactéria com a planta.

7 CONCLUSÕES

Os resultados obtidos neste trabalho permitem concluir que:

- 1) O genoma de *Herbaspirillum seropedicae* estirpe SmR1 contém moléculas de ncRNAs sendo transcritas em diferentes condições fisiológicas;
- 2) Os ncRNAs encontrados nessa bactéria estão distribuídos tanto no sentido senso como no sentido antisense do DNA;
- 3) De acordo com a região em que estão localizados podem ser classificadas em trans-sRNA, cis-sRNA e *Riboswitch*;
- 4) Utilizando dados de RNA-seq, em três condições experimentais distintas, foi possível confirmar a expressão de 98 ncRNAs;
- 5) Foi possível observar a expressão de oito ncRNAs que possuem alta similaridade com moléculas já descritas em outras espécies de bactérias;
- 6) A predição de alvos de mRNA permitiu identificar uma complexa rede de interação envolvendo o compartilhamento de alvos entre os ncRNAs.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABREU-GOODGER, C.; MERINO, E. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. **Nucleic Acids Res**, v. 33, n. Web Server issue, p. W690-2, Jul 1 2005.
- AIBA, H. Mechanism of RNA silencing by Hfq-binding small RNAs. **Curr Opin Microbiol**, v. 10, n. 2, p. 134-9, Apr 2007.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **J Mol Biol**, v. 215, n. 3, p. 403-10, Oct 5 1990.
- ALTUVIA, S. Identification of bacterial small non-coding RNAs: experimental approaches. **Curr Opin Microbiol**, v. 10, n. 3, p. 257-61, Jun 2007a.
- ALTUVIA, S. Identification of bacterial small non-coding RNAs: experimental approaches. **Curr Opin Microbiol**, v. 10, n. 3, p. 257-261, Jun 2007b.
- ARGAMAN, L. et al. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. **Curr Biol**, v. 11, n. 12, p. 941-50, Jun 26 2001.
- BABITZKE, P.; ROMEO, T. CsrB sRNA family: sequestration of RNA-binding regulatory proteins. **Curr Opin Microbiol**, v. 10, n. 2, p. 156-63, Apr 2007.
- BACKOFEN, R.; HESS, W. R. Computational prediction of sRNAs and their targets in bacteria. **RNA Biol**, v. 7, n. 1, p. 33-42, Jan-Feb 2010.
- BALDANI, J. I.; BALDANI, V. L. History on the biological nitrogen fixation research in graminaceous plants: special emphasis on the Brazilian experience. **Anais Da Academia Brasileira De Ciencias**, v. 77, n. 3, p. 549-79, Sep 2005.
- BALDANI, J. I. et al. Characterization of Herbaspirillum-Seropedicae Gen-Nov, Sp-Nov, a Root-Associated Nitrogen-Fixing Bacterium. **International Journal of Systematic Bacteriology**, v. 36, n. 1, p. 86-93, Jan 1986.
- BARQUIST, L. B., SARAH W.; GARDNER, PAUL P. Building non-coding RNA familie. **RNA Structure Determination**, 2012.

BARRICK, J. E.; BREAKER, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. **Genome Biol**, v. 8, n. 11, p. R239, 2007.

BENHAM, C. J.; BI, C. The analysis of stress-induced duplex destabilization in long genomic DNA sequences. **J Comput Biol**, v. 11, n. 4, p. 519-43, 2004.

BODDEY, R. M. et al. Biological Nitrogen-Fixation Associated with Sugar-Cane and Rice - Contributions and Prospects for Improvement. **Plant and Soil**, v. 174, n. 1-2, p. 195-209, Jul 1995.

BREAKER, R. R. Prospects for riboswitch discovery and analysis. **Mol Cell**, v. 43, n. 6, p. 867-79, Sep 16 2011.

BREAKER, R. R. Riboswitches and the RNA world. **Cold Spring Harb Perspect Biol**, v. 4, n. 2, Feb 2012.

BROWN, S. 4.5S RNA: does form predict function? **New Biol**, v. 3, n. 5, p. 430-8, May 1991.

CARVER, T. et al. BamView: viewing mapped read alignment data in the context of the reference sequence. **Bioinformatics**, v. 26, n. 5, p. 676-7, Mar 1 2010.

CHEN, S. et al. A bioinformatics based approach to discover small RNA genes in the Escherichia coli genome. **Biosystems**, v. 65, n. 2-3, p. 157-77, Mar-May 2002.

COENYE, T. et al. Identification of putative noncoding RNA genes in the Burkholderia cenocepacia J2315 genome. **FEMS Microbiol Lett**, v. 276, n. 1, p. 83-92, Nov 2007.

D'ALIA, D. et al. Noncoding RNA of glutamine synthetase I modulates antibiotic production in Streptomyces coelicolor A3(2). **J Bacteriol**, v. 192, n. 4, p. 1160-4, Feb 2010.

EDWARDS, A. L. Riboswitches: A Common RNA Regulatory Element. **Nature Education** v. 3(9):9, 2010.

EGGENHOFER, F. et al. RNApredator: fast accessibility-based prediction of sRNA targets. **Nucleic Acids Res**, v. 39, n. Web Server issue, p. W149-54, Jul 2011.

- FILIATRAULT, M. J. et al. Transcriptome analysis of *Pseudomonas syringae* identifies new genes, noncoding RNAs, and antisense activity. **J Bacteriol**, v. 192, n. 9, p. 2359-72, May 2010.
- GALVAO, C. W. et al. The recX gene product is involved in the SOS response in *Herbaspirillum seropedicae*. **Can J Microbiol**, v. 49, n. 2, p. 145-50, Feb 2003.
- GALVAO, C. W. et al. The RecX protein interacts with the RecA protein and modulates its activity in *Herbaspirillum seropedicae*. **Braz J Med Biol Res**, v. 45, n. 12, p. 1127-34, Dec 2012.
- GARDNER, P. P. et al. RNIE: genome-wide prediction of bacterial intrinsic terminators. **Nucleic Acids Res**, v. 39, n. 14, p. 5845-52, Aug 2011.
- GARDNER, P. P. et al. Rfam: Wikipedia, clans and the "decimal" release. **Nucleic Acids Res**, v. 39, n. Database issue, p. D141-5, Jan 2011.
- GEORG, J.; HESS, W. R. cis-antisense RNA, another level of gene regulation in bacteria. **Microbiol Mol Biol Rev**, v. 75, n. 2, p. 286-300, Jun 2011.
- GORDON, D.; ABAJIAN, C.; GREEN, P. Consed: a graphical tool for sequence finishing. **Genome Res**, v. 8, n. 3, p. 195-202, Mar 1998.
- GRIFFITHS-JONES, S. et al. Rfam: an RNA family database. **Nucleic Acids Res**, v. 31, n. 1, p. 439-41, Jan 1 2003.
- HERBIG, A.; NIESELT, K. nocoRNAC: characterization of non-coding RNAs in prokaryotes. **BMC Bioinformatics**, v. 12, p. 40, 2011.
- HOUSELEY, J. et al. A ncRNA modulates histone modification and mRNA induction in the yeast GAL gene cluster. **Mol Cell**, v. 32, n. 5, p. 685-95, Dec 5 2008.
- HUANG, H. Y. et al. sRNAMap: genomic maps for small non-coding RNAs, their regulators and their targets in microbial genomes. **Nucleic Acids Res**, v. 37, n. Database issue, p. D150-4, Jan 2009.
- JOHANSSON, J. RNA thermosensors in bacterial pathogens. **Contrib Microbiol**, v. 16, p. 150-60, 2009.

KEILER, K. C. Biology of trans-translation. **Annu Rev Microbiol**, v. 62, p. 133-51, 2008.

KINGSFORD, C. L.; AYANBULE, K.; SALZBERG, S. L. Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. **Genome Biol**, v. 8, n. 2, p. R22, 2007.

KOGENARU, S. et al. RNA-seq and microarray complement each other in transcriptome profiling. **BMC Genomics**, v. 13, p. 629, 2012.

KOZOMARA, A.; GRIFFITHS-JONES, S. miRBase: integrating microRNA annotation and deep-sequencing data. **Nucleic Acids Res**, v. 39, n. Database issue, p. D152-7, Jan 2011.

LANGENBERGER, D. et al. Identification and classification of small rnas in transcriptome sequence data. **Pac Symp Biocomput**, p. 80-7, 2010.

LASA, I. et al. Genome-wide antisense transcription drives mRNA processing in bacteria. **Proc Natl Acad Sci U S A**, v. 108, n. 50, p. 20172-7, Dec 13 2011.

LASA, I.; TOLEDO-ARANA, A.; GINGERAS, T. R. An effort to make sense of antisense transcription in bacteria. **RNA Biol**, v. 9, n. 8, p. 1039-44, Aug 1 2012.

LEASE, R. A.; BELFORT, M. A trans-acting RNA as a control switch in Escherichia coli: DsrA modulates function by forming alternative structures. **Proc Natl Acad Sci U S A**, v. 97, n. 18, p. 9919-24, Aug 29 2000.

LI, H. et al. The Sequence Alignment/Map format and SAMtools. **Bioinformatics**, v. 25, n. 16, p. 2078-9, Aug 15 2009.

LIOLIOU, E. et al. Global regulatory functions of the Staphylococcus aureus endoribonuclease III in gene expression. **PLoS Genet**, v. 8, n. 6, p. e1002782, Jun 2012.

LIU, J. M.; CAMILLI, A. A broadening world of bacterial small RNAs. **Curr Opin Microbiol**, v. 13, n. 1, p. 18-23, Feb 2010.

MAGALHAES CRUZ, L. et al. 16S ribosomal DNA characterization of nitrogen-fixing bacteria isolated from banana (*Musa spp.*) and pineapple (*Ananas comosus* (L.) Merril). **Appl Environ Microbiol**, v. 67, n. 5, p. 2375-9, May 2001.

MAJDALANI, N. et al. Regulation of RpoS by a novel small RNA: the characterization of RprA. **Mol Microbiol**, v. 39, n. 5, p. 1382-94, Mar 2001.

MARDIS, E. R. Next-generation DNA sequencing methods. **Annu Rev Genomics Hum Genet**, v. 9, p. 387-402, 2008.

METZKER, M. L. Applications of Next-Generation Sequencing Sequencing Technologies - the Next Generation. **Nature Reviews Genetics**, v. 11, n. 1, p. 31-46, Jan 2010.

MIRANDA-RIOS, J.; NAVARRO, M.; SOBERON, M. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. **Proc Natl Acad Sci U S A**, v. 98, n. 17, p. 9736-41, Aug 14 2001.

MOLL, I. et al. Coincident Hfq binding and RNase E cleavage sites on mRNA and small regulatory RNAs. **RNA**, v. 9, n. 11, p. 1308-14, Nov 2003.

MORTAZAVI, A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods**, v. 5, n. 7, p. 621-8, Jul 2008.

NIELSEN, H.; JOHANSEN, S. D. Group I introns: Moving in new directions. **RNA Biol**, v. 6, n. 4, p. 375-83, Sep-Oct 2009.

OTT, A. et al. NAPP: the Nucleic Acid Phylogenetic Profile Database. **Nucleic Acids Res**, v. 40, n. Database issue, p. D205-9, Jan 2012.

PAPENFORT, K.; VOGEL, J. Multiple target regulation by small noncoding RNAs rewires gene expression at the post-transcriptional level. **Res Microbiol**, v. 160, n. 4, p. 278-87, May 2009.

PEDROSA, F. O. et al. Genome of *Herbaspirillum seropedicae* strain SmR1, a specialized diazotrophic endophyte of tropical grasses. **PLoS Genet**, v. 7, n. 5, p. e1002064, May 2011.

PETERS, J. M. et al. Rho directs widespread termination of intragenic and stable RNA transcription. **Proc Natl Acad Sci U S A**, v. 106, n. 36, p. 15406-11, Sep 8 2009.

PICHON, C.; FELDEN, B. Small RNA gene identification and mRNA target predictions in bacteria. **Bioinformatics**, v. 24, n. 24, p. 2807-13, Dec 15 2008.

PIMENTEL, J. P. et al. Dinitrogen Fixation and Infection of Grass Leaves by *Pseudomonas-Rubrisubalbicans* and *Herbaspirillum-Seropedicae*. **Plant and Soil**, v. 137, n. 1, p. 61-65, Nov 1991.

RAGHAVAN, R.; GROISMAN, E. A.; OCHMAN, H. Genome-wide detection of novel regulatory RNAs in *E. coli*. **Genome Res**, v. 21, n. 9, p. 1487-97, Sep 2011.

REESE, M. G. Application of a time-delay neural network to promoter annotation in the *Drosophila melanogaster* genome. **Comput Chem**, v. 26, n. 1, p. 51-6, Dec 2001.

REPOILA, F.; DARFEUILLE, F. Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. **Biol Cell**, v. 101, n. 2, p. 117-31, Feb 2009.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: the European Molecular Biology Open Software Suite. **Trends Genet**, v. 16, n. 6, p. 276-7, Jun 2000.

RIVAS, E.; EDDY, S. R. Noncoding RNA gene detection using comparative sequence analysis. **BMC Bioinformatics**, v. 2, p. 8, 2001.

RIVAS, E. et al. Computational identification of noncoding RNAs in *E. coli* by comparative genomics. **Curr Biol**, v. 11, n. 17, p. 1369-73, Sep 4 2001.

RONCATO-MACCARI, L. D. B. et al. Endophytic *Herbaspirillum seropedicae* expresses *nif* genes in gramineous plants. **Fems Microbiology Ecology**, v. 45, n. 1, p. 39-47, Jul 1 2003.

RUMBLE, S. M. et al. SHRiMP: accurate mapping of short color-space reads. **PLoS Comput Biol**, v. 5, n. 5, p. e1000386, May 2009.

SCHMID, M. B., J.I.; HARTMANN, A. **The Genus *Herbaspirillum***. The Prokaryotes. New York: Springer: 141-150 p. 2006.

SELINGER, D. W. et al. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. **Nat Biotechnol**, v. 18, n. 12, p. 1262-8, Dec 2000.

SHARMA, C. M.; VOGEL, J. Experimental approaches for the discovery and characterization of regulatory small RNA. **Curr Opin Microbiol**, v. 12, n. 5, p. 536-46, Oct 2009.

SMITH, A. M. et al. Riboswitch RNAs: regulation of gene expression by direct monitoring of a physiological signal. **RNA Biol**, v. 7, n. 1, p. 104-10, Jan-Feb 2010.

STORZ, G.; ALTUVIA, S.; WASSARMAN, K. M. An abundance of RNA regulators. **Annu Rev Biochem**, v. 74, p. 199-217, 2005.

STORZ, G.; OPDYKE, J. A.; ZHANG, A. Controlling mRNA stability and translation with small, noncoding RNAs. **Curr Opin Microbiol**, v. 7, n. 2, p. 140-4, Apr 2004.

STORZ, G.; VOGEL, J.; WASSARMAN, K. M. Regulation by small RNAs in bacteria: expanding frontiers. **Mol Cell**, v. 43, n. 6, p. 880-91, Sep 16 2011.

TADRA-SFEIR, M. Z. Caracterização funcional de genes de *Herbaspirillum seropedicae* regulados pelo flavonóide naringenina. **TESE DE MESTRADO**, 2011.

THOMASON, M. K.; STORZ, G. Bacterial antisense RNAs: how many are there, and what are they doing? **Annual Review of Genetics, Vol 44**, v. 44, p. 167-88, 2010a.

THOMASON, M. K.; STORZ, G. Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? **Annual Review of Genetics, Vol 44**, v. 44, p. 167-188, 2010b.

TJADEN, B. TargetRNA: a tool for predicting targets of small RNA action in bacteria. **Nucleic Acids Res**, v. 36, n. Web Server issue, p. W109-13, Jul 1 2008.

TOPP, S.; GALLIVAN, J. P. Emerging applications of riboswitches in chemical biology. **ACS Chem Biol**, v. 5, n. 1, p. 139-48, Jan 15 2010.

TRAPNELL, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. **Nat Protoc**, v. 7, n. 3, p. 562-78, 2012.

TRAPNELL, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nat Biotechnol**, v. 28, n. 5, p. 511-5, May 2010.

ULVE, V. M. et al. Identification of chromosomal alpha-proteobacterial small RNAs by comparative genome analysis and detection in *Sinorhizobium meliloti* strain 1021. **BMC Genomics**, v. 8, p. 467, 2007.

VALVERDE, C. et al. Prediction of Sinorhizobium meliloti sRNA genes and experimental detection in strain 2011. **BMC Genomics**, v. 9, p. 416, 2008.

VAN VLIET, A. H. Next generation sequencing of microbial transcriptomes: challenges and opportunities. **FEMS Microbiol Lett**, v. 302, n. 1, p. 1-7, Jan 2010.

VERCRUYSSSE, M. et al. Genome-wide detection of predicted non-coding RNAs in Rhizobium etli expressed during free-living and host-associated growth using a high-resolution tiling array. **BMC Genomics**, v. 11, p. 53, 2010.

VOGEL, J.; WAGNER, E. G. Target identification of small noncoding RNAs in bacteria. **Curr Opin Microbiol**, v. 10, n. 3, p. 262-70, Jun 2007.

WACHSMUTH, M. et al. De novo design of a synthetic riboswitch that regulates transcription termination. **Nucleic Acids Res**, Dec 28 2012.

WAGNER, E. G. H.; ALTUVIA, S.; ROMBY, P. Antisense RNAs in bacteria and their genetic elements. **Homology Effects**, v. 46, p. 361-398, 2002.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet**, v. 10, n. 1, p. 57-63, Jan 2009.

WASSARMAN, K. M. et al. Identification of novel small RNAs using comparative genomics and microarrays. **Genes & Development**, v. 15, n. 13, p. 1637-1651, Jul 1 2001.

WASSARMAN, K. M.; ZHANG, A.; STORZ, G. Small RNAs in Escherichia coli. **Trends Microbiol**, v. 7, n. 1, p. 37-45, Jan 1999.

WATERS, L. S.; STORZ, G. Regulatory RNAs in bacteria. **Cell**, v. 136, n. 4, p. 615-28, Feb 20 2009.

WILLEMANN, G. M. PREDIÇÃO DE RNAs CURTOS EM *Herbaspirillum seropedicae* SmR1. **TESE DE MESTRADO**, 2010.

WILLKOMM, D. K.; HARTMANN, R. K. 6S RNA - an ancient regulator of bacterial RNA polymerase rediscovered. **Biol Chem**, v. 386, n. 12, p. 1273-7, Dec 2005.

YODER-HIMES, D. R. et al. Mapping the Burkholderia cenocepacia niche response via high-throughput sequencing. **Proc Natl Acad Sci U S A**, v. 106, n. 10, p. 3976-81, Mar 10 2009.

ZHANG, A. et al. Global analysis of small RNA and mRNA targets of Hfq. **Mol Microbiol**, v. 50, n. 4, p. 1111-24, Nov 2003.

APÊNDICE I – SEQUÊNCIA DOS ncRNAs CONFIRMADOS EXPERIMENTALMENTE

>Hs_noco_471 Hs_noco_471 undefined product 219112:219229 forward
CCAGGCAGTGAGATGGGCAGGTCAGTAGTAGATGAAAGCGTTGCTGGCGCCGAGGGTGTA
GGCCGGCAGGTTGTAGCGGTTGGGCGTGTGACGAAGCGTTGCCGGTATCGCCGGCA

>Hs_noco_494 Hs_noco_494 undefined product 457262:457311 forward
AGTGCCGCATAGACTGATGAAATTTTTGTCAGATCTCGTATAAAAGCGAG

>Hs_noco_525 Hs_noco_525 undefined product 932018:932087 reverse
GCGAGCGAGTCCCGTTTGGGATGAAGTGCAAGGCGCAATTTGGGTCAAGACTGGGCGTC
TTGACCCAAA

>Hs_noco_530 Hs_noco_530 undefined product 957549:957626 forward
GCGAAGAGTCATCGACGCGTCGGGAAGAAAGAACGAGCCAGTGACACCATCGGACAGCAAGGAATTCGGCA
TCGCCGG

>Hs_noco_534 Hs_noco_534 undefined product 1067480:1067548 reverse
CCGACCCACCGGTCCACCCAACATCACATCAAACGAAGACGCCATCCCGACACCGGCC
GGGATGGCG

>Hs_noco_637 Hs_noco_637 undefined product 2815035:2815179 reverse
AGGCGAAACGTATCGTATCGGTCTGGGAAGGCGTGCTTTTGTATAAGAACAAGCAGTTC
ATCGCCTTGCGCGGCGCTGCCTTGGTTGCCGAAAATGCGAGCCGTACATCCACGCCCTT
CTTCTCGCCTGCGCTCGCGCAGGCG

>Hs_noco_651 Hs_noco_651 undefined product 3173522:3173679 reverse
AGGCGGCGGTTTTCTATATGAATCGCTGCAACGGAACCTTGCCGTGGAAGTGAATCT
AGAACCCATTTCAATAATAGGCGTGAGTGCGAGCGAGTCCCGTTTGGGATGAAGTGAAG
GCGCGAATTTGGGTCAAGACTGGGCGTCTTGACCCAAA

>Hs_noco_682 Hs_noco_682 undefined product 3633077:3633376 reverse
TTGAAGGTAGCAAGTAACAAGGAAGTAGGGGTTGTACCGTAGCACGCAGTAGAAGTGGC
AATGATGCAGTAGTGGCAGGAATTGTAGAAGTTGCAGTTTTGCAGTGAAGCAGTCTGTGC
AGGAGAAGTGTGCAGTTGCAGGTGCAGTACCGTAGCAGAGCAGCAAGTGCAGTGGCAGTA
GCAGTAAAAATCTCTTGAGGGTGGAAACGGCAAGTCGTGAGGTCGCTTTTGTCCAAAGCG
CGCGTCACGTGTGGGAGCGGGTAGTCGACGCCCCGTGAAGCGGCCAGGCGGCTTGCCCC

>Hs_noco_698 Hs_noco_698 undefined product 3779834:3779991 reverse
CCGCTCTGGTTCTGCTGGCCGTACCCTGGGTGCGCTGGCTGGTTGCAACACCGTCAACG
GCTTCGGCAAGGATGTGCAAAAAGGTGGCGAGAAGATGCAGGACGCCAGCAAGAAGTAAGCAGTAAATTGC
AACAGGCTGCGGCCAGCGTGCCGAGC

>Hs_noco_700 Hs_noco_700 undefined product 3791718:3791799 reverse
AAGCGCTGCGCCACAAGGCCGGCGTTTCCGGCAGGCTCCTGAGGGGGAGGTTGGGACGGACGGCCCCGGTTG
GCGTCCGTCCC

>Hs_noco_709 Hs_noco_709 undefined product 3928375:3928556 forward
GCATTGCAATAACCAATGCGCAGGCGAATCGGACGTGAAATCTGTCCTGCAGAGTTTCCA

GCCGTATCCGCCCAAGCAGCTTCGCAAGCCTTACCTGTCATAACGACAAGTTCACAGTAA
 AAGCCCCCGCCCCTGGCATCCCTTGATGATGCCGTCCCGCGCAAAGCTTTTGTGCCGA
 AA

>Hs_noco_719 Hs_noco_719 undefined product 4074941:4075009 reverse
 AGGCCGTCAAATGCCTCCTTGGCGTGGGCAGTGAAGTCTTCTTCACTGCACCACGCCCA
 GAGAGAACA

>Hs_noco_722 Hs_noco_722 undefined product 4105880:4106014 reverse
 GGCCAGCTGGGCTTCGGCGGTTTCCAGTCTTCTGGTGCAGCTTGAGCATGCCATACAG
 GTAGTTGGCTTCGAAGTGCCGGGGCAGCTTACCAGCACCTTCTGTAGAGTTCTTCGGC
 CGGGGCCAGTTGGCC

>Hs_noco_723 Hs_noco_723 undefined product 4127435:4127552 reverse
 GTGCATTGCGAAATGACGGCTGCCGGATTGTGGTGCAGAATCGAGTCCGCTGTACGAAAA
 ATCCTATCTCGAAGAGGATAGAGGGGACACCAGCCCACAACGCGCAAGCCTTGCGGGC

>Hs_noco_726 Hs_noco_726 undefined product 4148338:4148890 reverse
 AGTTTGAAATGAACCGAAATGATGCTTACCGGCGGACTGAATCGCCGCGCGCTGGC
 TATGATTTGCACACGGGGAGAACAACCTGGAGGGCAGTCGACAGTTGTTGTCAACGTGGCG
 CAAAAAGTTTTCGTTAAGGATGCACAAGGCCGCGCCAGTGAGCACGCCGGTGGACTTGT
 GGAAGCACACAACACTGAGAGATAGAGGAAAACGATGAAGATCAACAACTGCTGGCTGC
 CGTACTCGTCGCCGCTTCGCCATGCCGGTCATGGCACAGCCCCGACCTGCCCGGGCTCC
 CGCCGCCGAAGCACCGGCCCAAGCCCAGCCCAAGAAGCACAGAAGAAGCAGAAGA
 AGCACAAGAAGCACGCCAAGAAGGCCGCGCCTTCGATGCCGATCAACAAGGGCGCCTGAGTCCGGGTCTCC
 TGCCAACCGTTTACCGAAGGGACGGCTGATACTGTTAAGGGATGCGCGCGCAGGGTATCGCGCAGCGCCA
 TCAGCCGAACCTGTAACCCGTCGGGTACCCCGCGGG

>Hs_noco_732 Hs_noco_732 undefined product 4239106:4239615 forward
 CCGCACTGGCGGGCGCATCTTTCGAGATGCAAGCGTAATACGCAAGTTCATGCGCCCAGC
 GGGATTTGCGAAGCTTCAACACCACGGGGCCGACTTGGTTTCGACAGGGGTTGCAAAGCG
 GAACAGGGCATGCCGAGGACTGGTTACCTCGTTAATCCATCCGGAAGTCTTAAATGCAA
 ACGACGAATCTTACGCTCTCGCAGCGTAAATCGCACCTAGAAACCTAAGAACTTCTAGC
 TCTACAGCACTTCTTCCCTGGGGTGTGCTGGCAACAGCGGTAGAGTCATTTACAGGGAAAT
 CGTGGCGTATCGGGTTACTTGATACGTCATTAATCCAAGGTAAGTCTGCTCCTGTCGAAGC
 GTGTGCGTCCGCGTCGCCAGGATTAACCAATTGGCTGCACTACGCATGTAGAAGTGC
 TGTAGAGTGCTTCTGGACGCGGGTTCAATTCGCGCGGCTCCACCAACATCACGACCACC
 AGCAACCGGCCAAGCAGGTTTCTGGTGG

>Hs_noco_742 Hs_noco_742 undefined product 4396086:4396487 forward
 GATCAGGATCACCTGGGGAATCGTCTCGCGCAGCGTGGCGCGCGTTCATCTGCGGCAT
 CAGATCGAAACGCGATGTAGATGAAGCCGAGGAGGCAAAGACCAGCACGAAGGGAATCCACTGCATCC
 CGCGCTCCAGCATGAAATAGCCCAGCACGCCCGCCGCCACGGCCATGAG
 GCTGCAGATGAGATTGTAGACATAGGCGCGGGTGCAGGAAACCGGCATTGAGCAACAC
 GATGAAGTCGCCGATCTCCTGCGGGATCTCGTGGGCGATGATGGCCAGGCCGGTCAACAG
 GCCAGGTGCGGATCGGCCAGGAAGGCAGCGCGGATGAGGATGCCATCGGTGAAATTGTG
 CAGGCCGTCGCCACCAGGATCATCCAGCCGGCCTTGCCGGC

>Hs_noco_744 Hs_noco_744 undefined product 4417512:4417615 reverse
 GGCCGGCTGGCTACGATGGCGGGCGCTGAATCGGACAGGCACGAATTTCCAGCAGGCGCAG
 GAGGCGTCGGGCTGGCATGGCGCGCCTCCGTGCCATGCGCCGCC

>Hs_noco_746 Hs_noco_746 undefined product 4428208:4428599 reverse
 CAGCACTGGCAGCGTCAGCACGGAGGCTACCTGATCGCATTCCGGTCGGCAGGTCGTGCGT
 ATCCTTCACTGATCGATTGCCGTCAACGAGGCAGGGGGGCACTGCCTCGGCGATTACAGGT
 CCTGTTTTCGGGCTTGCTTTATCGAAAGAGATGTCTCAGCCAGCCGCACCCGGTCAGTGG
 TGCCGATGGGTCGCAGGTCGTGGTTCGAAATCGACGCTGGGATTGTGCTGGTTGCACGGTC
 CCAGCCCATGGCCGGCTCGGCGTCGGCGCCGAACCTCCGGTCAGTACCTCCACCGCGAA
 GGTGGTTTCACTGGCGTGATTGCGACCAACTGCCAACTGCGCGCTGTGCAATTTCTGTTCA
 ATACAGCGTTCCTGTGCCTGTCTAGCATGCGC

>Hs_noco_748 Hs_noco_748 undefined product 4467241:4467750 forward
 CAGCAGATATTCGTGGATGTCGATGATGCGGCAGCCGGGGAAGATCTTCTCGAATTCATA
 ACCCTGCAACTGGTTCGTAGCAGGTACCGCAGGACACCACCGGTTTTGATGTGCGAGGTA
 GTTCAGCGTATTGGCCATGCGATGGAACAGCACGCGTTGTGCGGTGATGATCTTCTCGCC
 CTTTTCGAAGTCGCCGGTCCC GCGTTGCGGATAGCCGCAGCACAGGTAGCCCCGGCGGCAG
 CACGGTCTGTACGCCGACATTCCACAGCATGGCTTGCGTGGCCAGACCGACTTGCAGAGAA
 CAGGCGCTCCGAACCGCAGCCCGGGAAGTAGAACACCGCTTCGTATCGGCCGTGGTGGT
 CTTGGGATTGCGGATGATGGGGACGATCTTGTCTCCTCGATATCCAGCAAGGCGCGCGC
 GGTTCTTGGGCAGGTTGCCCGCATCTTCTTGTGATGAAGTGGATCACCTGCTCCTT
 CACCGCGGGCTTGCCGACGGTGGCGGGCGG

>Hs_noco_749 Hs_noco_749 undefined product 4473739:4474078 reverse
 CGGCCGGCACCATGAAGGTGGCGATGGCGATCTGTGGCTTGCCGCGCAGGTTACGCCAGG
 AAATCTCGTAGGCCGTCCAGAGATCCACGCCGAAGAAGGGCAGGGTGCCGCTGATCTGCA
 ATTGCGCGCGCTTGGGTGCGCGCGAATCGGAAACAGCAGGCTGGGGTCGTAATGGGTCCG
 GATATTGCGAACTTGTCCCAGCGGCGAGCTCTCTGGCGAATGGCTGGCGGGGGATGCGG
 ACGGGTTGGTGTGGTCCATGTGGATACTCCTGCGGTGCTGCTGATGACGTTACAGTGATGC
 TGCCGCGGCGCGCCGCACTGGGGTGGCGCGCCGCGTCCG

>Hs_noco_750 Hs_noco_750 undefined product 4474802:4474942 forward
 CGGACTGCCGCCGTGCCGCATATTTGGTAACATCTTGCCCTTCTGCAATATGAAGGATAA
 CGGGAAACTTTTGGCACAATCCCTGTCAATCCTGCATGGCCCATGCATGAGTTTGCCTGG
 CCGTGCGATGTCCACCATCG

>Hs_noco_761 Hs_noco_761 undefined product 4618621:4618928 forward
 GAAAGTCAGGCTGTAAAAAATATTGGGGGATATACCATGAGCAGGCCAACCATCTGGGAG
 GACAAGCGCCCCGCGCGGCGAAATCAGCATCGGAGGTTACGGCTAGGTCGCCACGACTGTC
 TATGCAGAGCAGCACATCACGACAGGGGTAGGGCTACAATAATAATAAAAAACACAT
 AAAGCAAACGCTGATGAGCAGACAGATAATGCGGAGCCTGTTGAGCGCCGCAATCTTCTT
 GGTCATTGGGTGATATTCATCGCCTACTGGTTCGGCATCAATCACCGGCTGGATCGCGG
 TGACGCCA

>Hs_noco_802 Hs_noco_802 undefined product 5239869:5240191 forward
 GCGGGGAGGGCCGCTCAGGAATTCCTGAGCGGGCCGCCGCTTTCTAGGGAAAACTACT
 AAAGCCAACGCCGGGAGATTTCTAAGATGGAATCATCTTGAGGAGATCACCATGTTCA
 CCACCTTCTGTTGACCGTCTGTCTGTCTGCTGATGACTGTGCTGCTTTCTTCGCTGG
 GGATCTATCTGTTGCTGAAGCGAAATGGCGTCATCACCCGCTGGCCTGCGATACCCGCC
 TGCTTGACAGCGCTGCCGATCTGATGTCTCCCTGAATCCTGGTGGATTGTTGGTGGC
 CCCACGGTCTTTGACCGCGGGG

>Hs_noco_820 Hs_noco_820 undefined product 5397591:5398115 forward

GCTGGCGAGACCAGGACCGCATAGCCATCCACGGCGCGCAGGGCGCAGTTCTCCAGGGCG
 GCGGGCGCACCCACGGCGCGCAATACCTCCGAGAGCAGCGGGCCGCGCAGGCGGTGGCGGGCCGTCGTA
 TTCCAGCGTGGGCTGGATCTCCACCGAGGGCAAGGCCGCCAGGGCGGCGAAGTCGAAGGTATAGGCCTTGTC
 GAAACTGAGCTTCTGCTTGCCATCATCTGGTCCAGC
 GCCGGATCGAGCGCGCCGCGATTGGCCTTGCCGATGTGCGCCGGTGATGGTGAGCAGGACC
 GGGCCAGCGCTGGCGCCGGCCTTCTTGTCAGCCGCCAGCGCCGGCAGAGCGACGGGCGCG
 GCCAGCAGCGAAGCGCCAGGCTGAGGAACCTGCGTTTTTGCATCGGATTCCTCTCTCTT
 CGGTATGGATGTGAGCTGTATTGATGGCGGATTGTAGACGCATCGGCGCGCAACCGACAA
 CGCACAAAGGGACAACGGCCCGGTTCCATCAGGAAGCGGGCCGTTG

>Hs_sRNA_4.2 Hs_sRNA_4.2 undefined product 2234033:2234214 forward
 CGCCGTCCAGGCTGTGCGACCGCGTTGGGCGGCCCAATCAGGCAATCCCTACACCGAA
 TCTGTAAGAGCGAGAAAGGCAAACCGTTATGACACCGCGAACTACCACCGAGTTTTCA
 CGCGACGCTAGTCGGGTTCTTTTTGTCCTTTGAAAAGCGCGGCTGGTCCGCGTTTTTT
 TT

>smallRNA_19.2 smallRNA_19.2 undefined product 612014:612436 reverse
 TGCAACGACGGTCGCGCATAGAACTCATTTCATAAATAGGCGTGAGATGCGTTCTGCCCA
 GAAAGGGCGCTGGCAAGGCGCGACGCGTGTGGCGGTGCCACACGCAAGGAGCGCAACGCGGCCAGC
 GCCTTTTCTGGGCAAGACCCGGAGGGAGCGGCCCGTTTGGGCGAATTGCTGCGTTACGAATTTGGGTCAAGA
 CGCCAGTCTTGACCCAAATTCGCGCCTTGCACTTCA
 TCCCAAACGGGACTCGCTCGCACTCACGCCTATTTATGAAATGAGTTCTAGGCTCTTCTT
 CAGCGACTCTGGCAAGACTGACAGTAGTCGCCACCCCTCAGTCGAGCGCTGCACCGATAT
 GGCTGGCGCAAGGCTATCCCTAAGATGAGGACTCGTAGGGCAGTGACGGCCCTGCGGTTT
 TTT

>4.5S 4.5S undefined product 4108897:4109042 reverse
 TGGCCCGTCCGGTGTGATTTGAGATACAATTCACCCCGCGGGCCCTGCGCATTGTGGC
 ATGGTCAACCTGGTCAGGTCGGGAACGAAGCAGCCACAGCCATTTCCCGCAAGTGCCGCA
 GACAAGGCTCGCCTCCTCCGTTTTT

>TPP TPP undefined product 462470:462762 forward
 TCAACCCATCCGAAATGAAACCTTGTTCCTCACGCCAATTGGAGTAGAATCCGGCCACG
 ACAAGGACTGCGCAAGCAGCCAGCGTCGCACCGCTAGGGGTCCTGGCGCTCAAGCTCATG
 CTTGGGTAGCCGGGTGAGAAATACCCTTGAACCTGATCTGGATAATGCCAGCGAAGGGAA
 GCAACCGGACGCTTGACGTACCCACGGCCCTGTATCACCTCCACGCCGCCGTCCTCTTC
 CGCAACCTCCTTTGCTGGCTCCGAGCCATGAGTCATCAAGCAAAGGAGCCGAA

>Hs_sRNA_Gs10 Hs_sRNA_Gs10 undefined product 808782:808981 reverse
 AATTGTTTTTCGTAAGTTGAACTTAATGTTGCTCAAACGTATCGCATAGACAGAACAGTA
 GCGGTCCAGTTCGCAATATGTA AAAACTTTCTAAATCGAGTGCATCAGGCACCTCCCGC
 TTACACAAATAAAACGCACGGGGCAGAGTTGAAAAGACCATCGATTACTAACAGCCTGGC
 CATCGGGACTGTTATTTTTT

>Hs_sRNA_Gs12 Hs_sRNA_Gs12 undefined product 2233545:2233793 reverse
 GCCGGACGCCAAGAACTGCATCGAGCAACATCCCAACGGACCGTCATCAACAACCGTGA
 ATACACGCTGCAAAGCCGTACGGTGCAGCTGGACAACGGCGAGCGGCACGAGGAGTACCG
 GGTGCTGCTGGATGGCCTGGAATCAAATCATGGACGCGCGGCGACGTCGTCCGTTATTT
 CGGCCTGGGCTGACAGCGCCCTGCTCCCGCACCATCAAAAACGGCATCCGCAGATGCCG
 TTTCTTTTT

>Hs_sRNA_Gs36 Hs_sRNA_Gs36 undefined product 1788675:1788860 reverse
 AACTAAGGCCGTCATGGCATTGACAGCTTTTCAATAGTCCTTTTCATTCCCCCACT
 GTAACGGACCAACATGATCAAGCCCCCTTTACCCGCTGCTGGCGCTGACGGCGCAGA
 CCGTTGCGCCTGGCGCGGCTGCTCTCGAGGCCGGCGCGCAGCCACGCCGACCTCGACT
 TCTTCT

>Hs_sRNA_8.2 Hs_sRNA_8.2 undefined product 5335529:5335812 forward
 GAAGAACCTGCACGACCGATCGGCATCGGTCTTGTGGTGCAGAAAACGTCAGGCGCGGT
 GCAGGAAACATCAGGCTTCCAGCTCGCAGGAAGTGACCGGAGGGCCTGCACAAAAGCAAGAAAATTAGA
 AAACCAGACCGTCGTGGCAGACCGGTGAAACACTCCGGCAGCACGGGAGAACGGCAAACAAACTCGGA
 GACACGCTCGCAAGAGCGGCAGGAGGCCATGATGGCTCACGCCATGCGGCGCAAAGGCGCGCTGCTGC
 AGCGTTTTT

>Hs_noco_498 Hs_noco_498 undefined product 551015:551235 forward
 GCTCACGGGTTTGTCCCGTGTAGCAAATAGTTAATTATGGTTTCGTACCAGCGTCGCTCC
 CCCGTCAACACAACGGCGGCAGGACAATAAAACGGACAACGTCATGACGACGGCACCGTC
 CAGATAGGCCGGGTACGTCAGGCCAACGGCACTGCGCAGATAGCGTGGGCTGGGCCAAT
 CTAGTGTAATCAATGAACGGCCCCGCTGATTCAAGGCGGGC

>smallRNA_49.2 smallRNA_49.2 undefined product 4896565:4896712 forward
 TTGGACCAGCTTCGGCACTGACGTTTAATAGACCTCACTTGAACAGCGATTGATTGGTAT
 CCGTTACTGTGCAAGGTTCTCCCTCTAGTGTTGAAAGCACCCTAGCTCGCGGATCTG
 GCTCCCCCTGTGAGATCCGCTTTTTTTT

>Hs_noco_786 Hs_noco_786 undefined product 5019358:5019702 forward
 GTCTCAATATTTCACTGGCGCTCTAAAAAGTAATGAAATGTACGGTAACAACGATTACTC
 AACTTTCTATTACATTTAGCGCAGCGGAAGTTACGGCCGGGCATCCAGAGAACAGGCAA
 GACCCATGGTGAGGAATTGATATCGCGCAAGGTATCTATCGTCCAACACCTGCATACTGT
 TTCCAGCAGAATACGTTGCTGTTTTGCAATTTTGTATCGAGCCTGGCACTTGGTTGAT
 AGAGAAGACCGGCTCGCGCCTTGACCTGCATTGAAGGTGGACGCCCATTTATGCCACAG
 TGAATTGGAAGCATAAGGCTGCCAACGGCGGACTTATCCTTCG

>smallRNA_43.2 smallRNA_43.2 undefined product 4564951:4565134 reverse
 GTGACTCATGCGGGAGCATCCGGCGTATCGTAGTCCGGTAGTCCCATAGTCCCATCATGC
 AAGCAAGCCCTCAACCAAGACCAGACCGAAAAGGCGTCCGGCGCTAGGAGACGGATGGCGCGCCTGTGTTT
 TACCGGCTTTTTGTCTCTTGCATAGCAGCGATTGCAGTGGTTGCAGT
 TTTT

>Hs_noco_725 Hs_noco_725 undefined product 4136562:4136795 reverse
 CGATGAGGTTTTTATGAACGTTTTTGACGTTTTGGACTATTTTCGGCCATTCTGGAAT
 TGCTGCCGATACCTGAATCGGTTATAATGTGTTAGTTAAACGCAACCTGCCGCGAATCGC
 ATATGTGATAAATCATTGATAGATCCGCTGCGCCATGAACGATGGAAGTACGGTGGTTC
 GCGTGATGTGTGATCTCGCGTTGCGTGAGATAACAAAAGATGGGCAGATGCCCA

>Hs_sRNA_Gs1 Hs_sRNA_Gs1 undefined product 3998440:3998528 reverse
 TGGTAGCTTTTTAGAAAATTTCTTCAGCAAACGGAATCGTTCGCAAGCTTACGCATGGC
 AATGGCTGCCAGCCGGGCCGCTATTTTT

>Hs_noco_640 Hs_noco_640 undefined product 2939684:2939824 forward
 GCGTTTTGACACAAGGATCGCGGCCGCTGTGCGACTGCGGCTTACGCCAGGCGAG
 CCGGGGATTACAACATCAGCCATGCGCGATACCGGGAAACCGGAGGGGGTGGACGGTGT
 ACTTCTTGATGAAGTACACCG

>Hs_sRNA_45.2 Hs_sRNA_45.2 undefined product 2140728:2141053 reverse
 TCTGGCAAGACTCGCATCCGGCGTTCCGGCCCCAGGACAGCAGCGCACAGCCCACCGGCGA
 CCAGCAGCGCAACAGCGGCAGGCTCATCCGCCTCCCCTCTTGCAGCGCATCGCACGGGCC
 ATGCCGCAATTCAGGCCGAGGCGTCGCGAAGCTTGCCGGTTGATGCAGTGAAACCCGATC
 GCCGTGCGCAACCCTATTGATATGTTACGTTTGGCTTCACAGACCAGCTTACCGTCCGCA
 GTTGCTGCGCGACCGCCTATGGCGGAGCGCGCGGCGGATACGGCGCTGCCCGGCGTT
 CCGGTCCACCGCATCCGCGCCCCTT

>Hs_noco_564 Hs_noco_564 undefined product 1587943:1588410 forward
 GCAGCTTTGCAAGCCTGGACGCAGCAGCTACAATAGCGGCCTTGCTGCCCCGTAGCTCA
 GTGGATAGAGCGATCCCCTCCTAAAGACAATTGGGCAGCGGCTGCCGAAAGGCGCCGCGT
 GAATGGAGTCAAATTCGGTGAACCCTGACAGCGCCTGGTCCCTGTGACTGCGCGGCTGCT
 GGCAATACCGAGCCAAGCCGCTCCGAGATGGATGGAAACCGGAAGGTGTAGAGACTAG
 ACGGCTCCGGCCTGCAGGCATGTGCGGTATCAGGCACAGGCCCATGGCCAAGGTATAGTC
 CAGACCACAAACGTTCTCCGGTGGCGAAAGCCATAGTGGTATGAAGGGATAGGTCACACG
 TTCGATCCGTGTCGGGGGCGCCAGCATAAAAAAAGCGCAGAGGGCAACCTCTGCGCTTTT
 TCGTTTACGGCGTCTCCTTACGCTCTTCTCTGCATCAGGAAGAGGC

>Hs_noco_565 Hs_noco_565 undefined product 1596987:1597275 reverse
 CCGCTGCCGTTATTGACAGTGATTGCTCGAATCATCGTTTGCCTCACATCCCTAACATG
 TTGAATTGAAGGACGCAAAATGAAAAAATTGATCGCCGCTCTGATCGCTGGTCTGTTGCG
 AACTGCTGCTTTCGCACAAGCTCCGGCCGCTCCGGAAGCGCCGAAGGCTGAAGCCGCTGC
 TCCGAAGGCCAAGAAGGCAAAAGAAAGCCAAGAAGGCAAGAAGGCAAGAAGGCAGCTGAAGCTAAGTAAT
 TAGCGTCCGCATGACGAAAAAGGCAGGCTGGTCTCTGCC

>sRNA46 sRNA46 undefined product 1636668:1636831 forward
 TGCTGCTTCACTTAAACGATTGACGCCAGACATGTCCTTACCCGCGCTTACTTTTTATT
 TTTACTTTAGCTATTCCAGCCCCACGGCGGTGAAAAGCGGTAAGCGTTCGTA AAAAAGACA
 AAACCGATTTCTAAAAAACCGCCAGTGATGGCGGTTTTTTTTT

>Hs_noco_593 Hs_noco_593 undefined product 2032405:2032540 forward
 GCCGATGGTAGAATGACCTTGCCAAGATGGTAAGTTTTTCACTCCAGGATACCCCGCA
 GGTATCGCCCGCAAAGAAAAAGAATATGGGTTTGGCTAGCTGGATGACCGGATCGGTGGC
 GTGGGTGCCACCGGTC

>Hs_noco_551 Hs_noco_551 undefined product 1383591:1383913 reverse
 ACGTGCAGGAAAAACCAGAACCATTTCATCAATAGGTTCAATAGGTTCTAGTTACAAAA
 TCCTCGGGGAAAAATTACACAGGCCTTGATCGTGGCAGCGAATGCGTTAGACTTCGTA
 CTCCTGCCGTGTTTCGCGATTGCCATATATTCTTGAACCATTTACGAGCATAGGTTGTGG
 GATCGTTTGTATGGGCGTGAGCGTCGCTAGTCCGATGAACCCGAAATTGAACTACTGCAAC
 CAACTTGAACCGTCAGGTTGAGGATGCCGGCACAGCGGCACAGGCGGGGCTTGATTCAAG
 AAAACGGCTGCATATGCAGCCG

>Hs_sRNA_Gs25 Hs_sRNA_Gs25 undefined product 1433087:1433246 forward
 GCGCTGTATCGTAATACTGACGCCTCGTCAGTAAGTACAAATACGAATGCCTTGCCAAGG
 TGCAAGTAGGAACAATATCGCACCGCAGGAATCGGCGCCCGGCCACCGGAATCAACGGCC
 CTTTTCTTCAAGCTGTCTTCTGACACCAGCAGTTTTTTT

>Hs_sRNA_Gs23 Hs_sRNA_Gs23 undefined product 935267:935546 forward
 GCGGGCTTCGCTGAAACGGCAAACGGTCATTGCCGACCCAGCAATCGAATAGAGCAGTCT
 TGTTTTGGTTCAATTGTTGTTGATGTACGCGCCTGGCCGGCGACCCTTCCGAATCATCGG

AACGGCGCACCCGGGCGGCACCCGATACAAAGCGGGTGGTGCATAACAAACGGTATCGAT
 ACCGGGAACTGAGGAGACGTGCATTGATTGAAGCCGTTCAAACGGCCGGGAATGCATC
 AGGAGTATTTGAAGGCGCATCTGCGGATGCGCCTTTTT

>Hs_sRNA_57.2 Hs_sRNA_57.2 undefined product 417221:417409 forward
 GATGCCCGACTGCTGAAACGGAAAAGGCCTACGCCACATTAATGAACAGTCAGAGACCGA
 ATCAATCTTCTCTGTTATCGCCGCTTAACCCCATCTTGTCTGGGTGAGGCGGCTTTTCCTT
 TGAAAAACACCGAGGTTTTTCAGGAACGCGATTCCGACGCTAGCGTTTGAACACGCTGGC
 GTTTTTTTT

>Hs_sRNA_Gs11 Hs_sRNA_Gs11 undefined product 808938:809200 forward
 CGATACGTTTGAGCAACATTAAGTTCAACTTACGAAAAACAATTACTTAAAACTTAAA
 ACTTAATTAATTAGTTGGTTGCCAGCCTCATCAACCGGCTGGTCGCGATTGGAGCGCGGG
 ATGGATAAAAAGTGCATTTTTTCGCACAAATTTTCGCGCAACGGTCGCTCAGGGAATTCA
 TTGATGTCTCAACAGTAAGCGTTCGGAGATCTCCGGGTGCTGCATGGATCGTCTGCGCGC
 CCTGCGCGCAGATTTTTTTTTT

>Hs_sRNA_Gs4 Hs_sRNA_Gs4 undefined product 4127685:4127782 reverse
 AGCCATGATGCACTGCCATATAGTTACACCTGTCTCCTCCAATTCCTCCTAAAGAATTG
 GATTCAAGCCCCAACTTCGGTTGTGGGCTTTTTTTTTT

>CUFF.55 CUFF.55 undefined product 205713:205798 forward
 TTTGGACAGCTTGCCAGAAGTCTGGGACAGAGGCTGTCAAAGCATTGCCACGCCCCG
 TTTTCGAAGTTGTTTCGATGCTTTCTT

>CUFF.111 CUFF.111 undefined product 359683:359916 forward
 TTCCCGCAACGGTGTGTTAAGGAAGTGCAGTAAAGTCTTTGACTTTATTAGCAAAATCCCT
 GTTTTTCCCTGCGTATACCGCTTGACCTGCCATAGTGCATCCTCTAAAGTGGGCGGCAGT
 GTCAAAAAGTGTTTTTTTGTGGGGCGTTTTTGCAAAGTCATTTGCATTTGCCCTTCGCGA
 CAAGACACCCCGATCAACAATCCCAGTAATCTCACTCGCCACAGGCGAAGGAAA

>CUFF.124 CUFF.124 undefined product 412442:412676 forward
 AGTCCCAATCGATGTTTTGATACTAGATGTAGTGTCCAGGTGAACTTATTTTTGTACACG
 GGTTCTTATCCACAGATGGCCCGGCTGGCAGTGCTTCAGAAGGCGCAAGATCCCAGGAG
 CAGTCAACCAAGTACCAAGCAACAAGTACGAAAGCGCCCCGTTCTCCCTTCTATTCCGT
 CGTTTCGCTCATCAGTTTCAAGCAGTTTTTGCCTCAGCATGAATATCCATGCAAG

>CUFF.223 CUFF.223 undefined product 664741:664801 forward
 AGAACTTACTTCTTAAATAGGCATGAGTGCCTCCGTCAGCAAAAGGCTTCTCCACAGCC
 A

>CUFF.251 CUFF.251 undefined product 741007:741139 forward
 TCAAGAATAGTACGCACGTTCTTTTTGGTGGGAGGCGTGAGTTCAGGTGATGCAGAAGT
 TACTGATCATTACGCCGACAAGAAATAACGAAAAGCGGTCATCAGCCGTAGATAAGATC
 CCGTCAGACGCGA

>CUFF.273 CUFF.273 undefined product 792569:792670 forward
 ATTGAAACCAAGAATCAGTGGAGACAAGGAACAAGCGCAGGTTTCATGCGTTAGTTTATCG
 GCTGCCCAAAAACCACCTGATCCGGCGCCATCGACTGCAAT

>CUFF.395 CUFF.395 undefined product 1155828:1155922 forward
 CGGGCTGTACAGGACAAAACAATGACAGTGTCCGGAATGAGCGAAGAGATGGTCCGGGCG

CATCAAAACAAAAAGTTGGGAGACCGTAAGGGGTC

>CUFF.404 CUFF.404 undefined product 1189098:1189333 forward

ATGACCTGTAGAGATAGTCATTTAATCCAGTGTTGCCATTCGTGACGAGGTGAAGAATGC
AAACGAACCAAGTTTCCAAGACCGAATCAAGGCCAAGGCGCTCGAATACTTCCGCCAAG
TGAAGCCTCTGGTGAGACCCTGGTGGTGACGGATCATGGCAAGCCTGCCGTGGAGGTGC
GCCCATTTCCGGAGGCTCAACGGGCGCCCCTGGCAATCCTGGCTGGGTCCGTCCTG

>CUFF.504 CUFF.504 undefined product 1336439:1336564 forward

AAGCTGACACGACACAATCAAGGCCGGAGACTGCTTAAAAATCAGGCGTTTGTTCATGGT
TCGGTAAGATTTTGTGGTAAGTTTCGTCCCAGAGTTGATGCAAACGGCGCGCAAACAAGC
TGCATA

>CUFF.522 CUFF.522 undefined product 1381282:1381335 forward

GACCATTTCTCTCACGTGACTGGCGAAACGCCGGCAAAGCCGGCCAAGATGG

>CUFF.524 CUFF.524 undefined product 1382857:1382981 forward

GTTGTTGCGGGCGCCTTCGCGTATCTGCCTCTCGGTACACTGGATTGACCGGCGCGTTGG
TGTCATCGTGCCCTACCCCGGTGCCGCTGCCGGCAGGGCGGATCTGGTCAGGTCTGATCTG
ATCTG

>CUFF.689 CUFF.689 undefined product 1867577:1867679 forward

ACGGGATGGGAATGTACTCAGGACGAACGGACTGCGTGAGCTGAAGCGTGTTCAACTCCG
GTCTGACGATACCCAGCGCCGAGGAAATCTGGCTGGCCGTG

>CUFF.690 CUFF.690 undefined product 1880603:1880826 forward

CAAGGAGCGCAACCGCGCCAGCGCCCTTTCTGGGCAGAACCCGGAGGGAGCGGCCCGTTT
GGGCGAATTGCTGCGTTACGAGTGGGTCAAGACGCCAGTCTTGACCCAAATTCGCGCCT
TGCACTTCATCCCAAACGGGACTCGCTCGCACTCACGCCTATTGATGAAATGGGTTCTAG
CTACTTCTTACGCTGGCGGCGCATTGGTCGCGCAGGGTGGTGT

>CUFF.738 CUFF.738 undefined product 2011544:2011601 forward

GAAGCTGGGCAACAACGAGTAACGAGACGAGCCGGTGTGATATTGCCGGTCCGGGCGGC

>CUFF.839 CUFF.839 undefined product 2225999:2226140 forward

CGCCCGACGTCTTCTGTAGCCCGGAGCGCCCGTCTGCCTTTCAGGTGAGCCGGGGTCA
GCAGACCTGGTGTGATTCCGGGGCCGACGGTCATAGTCCGGATGAAAGAGAACGGGATGA
TCGGTCAACGCCGCTGCCCTGC

>CUFF.852 CUFF.852 undefined product 2287397:2287464 forward

CAATGCCGTGGCAACCGAGAGGGAAGCGAGCGCCAGGGCGCTTGCCATAACAGGTGGCG
GCTCAAGT

>CUFF.934 CUFF.934 undefined product 2519484:2519726 forward

CGGCGCTGCATTCAGGCGGGATGCACAAGGGGCCGGCAGTGGGGAATCCGCTTCACTTCT
TCCAGTTGTTCAATTGATGCCGGGCCACTGTTGCAGGCGGTGAGAACATTCAGTGTTCCT
GATCCGAGATGAAATTCCTGTATCCGCATGAAGGTGTGCGCCGTTCCGGAAAGGGAGG
GCTGCAGCAGCGTTGTGTCGACGGCACAGAGCAGTCCGATCCGGAAGGCGTTGCCCGGC
CAG

>CUFF.938 CUFF.938 undefined product 2526174:2526304 forward

ACCACGGTATAATGCGCGCTTTGCCAGCTGACGCGCGCACGATCCATCATTCCGACGACC

CCCGCCTGCACGGTTTTTCCCTTGTGCAATGCGCCATGACGGTTGGCCGGTGCCGGTG
GACCGATCCCT

>CUFF.967 CUFF.967 undefined product 2578050:2578397 forward
TGCTGATTTGTCAGTTTGAAGTGTACCATGCATTCCGCTCTGGAAGTAAAGCTCATCTT
TTGACCGTCAAAAAATAAGCCGGACGCCATACCGTCCGCGCGCTGGATATTGGACGTGAC
TCAGGAAGGGTGTAGTCCGCGTGCACGGCTGGCTCGCGTTTGCAGACCTGCTCCAGAC
CGCGCGCAAGGCCGATGAACCCAGCCACCTGATGTTTCTCCGAAATGCCACGGCATATG
CCGACGCTGCCACGAGGACACATATTCATCAAATAGGCCTTGCCTGCGCCCTTCTGTT
CGAATGCCAGGCGCCTGCCTTGTCTAATGTGTTCCGATAAAAAACGC

>CUFF.1011 CUFF.1011 undefined product 2800595:2800705 forward
CTACCATTGCGCCGAGGAGATGGCATTCTCCTCAAACCGCCCAAGTCCAGCTTGGG
GCTGATGATGCCTACAGTGACCTGGCAACCCGGGGCTGTAGGTATTCGCT

>CUFF.1094 CUFF.1094 undefined product 3160202:3160357 forward
CCATCAGCACGAAGAGCGCGTAGAGCATGGCGGTCAGGCTCAGGTGCTTGTAGACATACA
GGCCACGTAGAGCAGTCCACCGGATCCAGACCAGCCAGTTCTCGATCTTCTGCGCG
AGAGCAGGAAGTCCCCACCAGGCTGCCCGCCGTGA

>CUFF.1117 CUFF.1117 undefined product 3204507:3204585 forward
CTCAGGAGGAGACGAAACGGCGCATATCGCTCCGAGCTTTCAAGGATGCGAAATCAAT
GCGATCGTCATGATGCAGC

>CUFF.1142 CUFF.1142 undefined product 3303508:3303560 forward
GGGGTAGTACATGCTTATACGTTTAATCCATTTCTCATAGGTTTCGATCGCTT

>CUFF.1223 CUFF.1223 undefined product 3512060:3512227 forward
TGTGGAACCGGGAAGTCCACCAGATGCAATGCTTGCAGCATTTTCGTGCGACCCTGCCA
GTGCCCTGCAGAGGACCTCGCCCCGCGCATGCGCGCCGGGCTGGGTCATTCTCATCAGT
GATGACTGTTATCAGCAAAAACGCGGGTGCATGACATTGCGATGCGGC

>CUFF.1716 CUFF.1716 undefined product 4925909:4926013 forward
AGGGGATTTTCTGGGGTGAAGATCCATTCAATTGGGAACGAAGCCGCATGACGTTTATA
ATTCAGGGCGAGCAGCAGCCAGGCCCTTGCAGCGCCTGCGCCACG

>CUFF.598 CUFF.598 undefined product 3781409:3781584 forward
CTCCCAGAGTTTGTGCTAGAGCGTCTTGAGCATGTTGCGGCCTGGCGAAACTGAAAAGGG
ATGCTGGTCAGGAAAATAGCATTTGATTATGCCATAAAGTGATAGGTCTGACCTCCCTGC
CGTGTTTTCCGCCGCGCCGCAACAACAGTGAATAATGCTGCAAGGTCTCGAAAAA

>CUFF.225 CUFF.225 undefined product 539844:539917 reverse
AAACGTCAGGCGCGCCTATCAGGACTGCAAGGACTACGAGGACTGCGCGCCTGCATCTC
GGCAATCAGGGCGA

>CUFF.228 CUFF.228 undefined product 541704:541796 reverse
AGAGAATGACATTACATCTTTAGTTGTTGTTGCAAAAAGCAGTGGGTCTCTCTAGCACG
AGACTATGCCGATCAGCTAGTAACTGAGATGAA

>CUFF.1027 CUFF.1027 undefined product 3181155:3181240 reverse
CGTTTCGTCCCGTGCGCCGATGAGAACAAGAAAGCTGCTTGCCCTCCTGATCCCTTGAC
CGCCTTGCTGGCCGGCTGCGCCAGTT

>CUFF.1106 CUFF.1106 undefined product 3382844:3382966 reverse
GTGCGGAGCGACCCGTTCTAGTCTCCGGAGTCCATGGAGGAGGGCGGAAAGGCCGTGC
GGCAAGGCTTTGCGGCTGGGCGTTGGTGATAGGAAACGCCCGTCGCATCCGGTTTCAAAT
AAA

>CUFF.1109 CUFF.1109 undefined product 3382646:3382739 reverse
CCCACAATTGATGCAATCCGCTAACCGGTCAGGCCGTGTCGCGGAAGGTGTAGAAACCC
GCTAATCTCGGAAGCGCAAGAAAGGTGAGCAA

>CUFF.1158 CUFF.1158 undefined product 3553031:3553120 reverse
TCAAACCAAACACTACATAGTATTTATTTGCTTATATACGCCTTTGGCATATGGAAATGCGT
CAAACGAGGCAATCAGGCCGAGGAGGCCAG

>CUFF.1190 CUFF.1190 undefined product 3614107:3614194 reverse
AAGCAATTGCGGGCCTTTTCAATTAGTGTGAGGCTGGTAGTTTGGGTCTCGCGCAAGGG
CGGCAGGGCTATGCCGCCACGATGCAGA

>CUFF.1194 CUFF.1194 undefined product 3630870:3630996 reverse
TTCTCTTACGCCCACGCCGGCCAGTGGCCTTGTAACCCGGCAGTACGCACACGTGGTGG
AAGCTGCGGTCTCCCTCCCGACTGGCGCAAGCCTGGTCGGTGC GCGGCGCGGGGTGCG
TATCTCT

>CUFF.1510 CUFF.1510 undefined product 4402051:4402283 reverse
CATCAATCTCTTTGCTGTGTAGCAAAGATGCGGGTGTCTGGGTAGAATCTCAGCACGA
GCAGCACAAGCTGGGCAGTGCCGGAACACAGACGCGGGGGCGTCTCCGATCCTCTCCAG
CTTGCCCTATCGACACCGAGGCCTGCATTCCCGGATGAGCCGTATCGCCACGTCGCGAT
CGCGCCATTCCGGTGCATACGACGGGCCATGCATCCACAGTCCAAAGAAACA

>CUFF.1548 CUFF.1548 undefined product 4499669:4499780 reverse
TCTACAGGATATGAACAGGGTGTAAACATTTGTA AAAAGCGCTGCCTCGGACCGCTGATT
TTCTTATATTGAAGTTCCCTCCATTCTCCCCCGAGAATGGGATTTAC

>CUFF.1570 CUFF.1570 undefined product 4587035:4587149 reverse
AGGCGGACGCAGCAAATAGGGAAATAATTTATATCCATTACCAAGCGGCCACAAAAAC
AGAGCCGCAGGCCAGAAGGAGACAACCCGCATCATGCATCTGCATCATGAAGGCG

>CUFF.1692 CUFF.1692 undefined product 4834369:4834486 reverse
CCCTCGACATGACCCGATCCCGGTCCGCTCTGACGGAGACGATTTGTACAGGCATTCCCC
AAATGCCGGAGACAGGTATCGGACAGGCAGTTGAATCCGGCAGGCATGGCCAGGCAA

>CUFF.1713 CUFF.1713 undefined product 4909594:4909823 reverse
AGGAACACCAGCCAGCCACCACCGCCGATGGCATGCAAGGCTGGCAAGGGCAGGAAAGAGAGCAGGCG
GAATAAAGAAACGAGCATACTTGCCGATAACGTTGGGCGGGGTGCGTCCGCCCAAATTGCAGAAAATTG
CAAGAGGCGTAAAATAGCACAAGCCGTGAGTTAATG
ACAACCTGCGAGACGGGTACAAATATCGCTAAAGCGTCGCAGGCACGAC

>CUFF.1720 CUFF.1720 undefined product 4929103:4929157 reverse
CCCCCTGTAGTACCCGTCCCGACCGGCTCTCCCCGGGCAAGCGGCCGACCCTCA

>CUFF 17.51 CUFF 17.51 undefined product 5000211:5000493 reverse
GACTCTTTTTGATTAAATTGGCGCCTGTTTCAGAAAACGTTTACTCCAACCTGAGTATTTTT

TCATCTCCGTCCCGGGTTGGCCTGCGCCTGTCTCATCGGATGAGCGGATGAAAGCGGTGC
AGGTCAGGCGCGTAGCAGCCCCATTTTCTGCTGCACTGCAAAAAATCAAAGTGGCGGCTT
ATTTGCCACAAATGCCCCGCAAGTGCCTGAGATAAGGCCTTGATCTGGCTGCAAAAGGTT
TGACTTACCAGAAAACGAACCTTTATAATCGCCGGGCTTTTAA

>Hs_noco_652 Hs_noco_652 undefined product 3180163:3180284 reverse
CGCCAGCGCTTGC GCGCCCTGCTCGAAGGGTAGCCGAGGTCGGTCACCAGCAGCACGAA
TTCGTACCGCCGACCCGCGCCAGGGTGTGCGCCGTGCCGCGCCACCTGGCGCAGGCGGCG
GG

>Hs_noco_764 Hs_noco_764 undefined product 4643564:4643620 reverse
TCCGCGCAGTTTTTGC GGAATTGCGTATCGGGGTGAGCGTTATGCGTCTCGCCCCC

>Hs_noco_656 Hs_noco_656 undefined product 3201693:3202171 reverse
GTGTAAGAGGGGCTGTCAAGTGGATCTGAATTTTCGGGCGCAGGCAGCGTATCTGCCAGGGA
AGGTAGATCATCGTCGCCCCACGCCGCGGGTTGATCGATGCTGACGGCGTACCTAACCGA
TTTACCGGGGCGGACAGTCTGGTTGTAGACAGTGGCCAAGATCCAGTCTTGATCTTGCTT
GTTTCGAGAAGTCAATTTCTGACTTTCTTCTCTCCCAATTCTGCAGCGACGACCCACGC
GCTGCTCTGGAGCTCTCAAGTTCAACCTGTCCGCGAGCTTGTGCAACCATGCGGCGGAA
GTCTGCAATCTTCTCTGATAGAAAATTTTTGAACTGCTGGTACATCAGTTTGCCGAGGT
CTATGAGTTGCCCGCAGCATACTTTCTCCATGTTACAGGCGTATGAGCGCCTTGCTACC
ACGCTTTAGTTTGTGGGTGAATTTGTGGGTAGAAATGATAAACGGGCTGAAAGGCCCG

>Hs_noco_819 Hs_noco_819 undefined product 5373859:5374363 reverse
GGACGATAGCTGGGATTTCCGCTGGGGCGACCTAAGCGATCAAGCTGATCAGCGATGCTC
TCGCCGCCGAAGTGAATCCAATTGTTCCAATATGCTTCGTCGAAAAGGGAGCTCTGGTC
ACTTTATATGTCCTGAAATATGACTGACATGACTGACAACCAAGGAGCGAGCCTTGATGA
TGACCTCTGCTTGACTGCGATCACCCTGCTGGACGAGGATGGCTGTACAGGAATCTTG
TCACGGTACCGCACTGGTGCAACGAGAGAAACCTTCTTCAACGCAGCGCCGAGTTCTGC
ATCGCCGCTGTTGGTGCTGATAAAGATGATCTCCGGGCCATCATTGCTGATGCCGGAATA
GCCATCGATATTCTTGTTGGTGCCTGGGTTTATCTCGATTTTGGCCTTGCTGCCGTAAT
CCAGCAGCACGCCATCTTTCTCATCATGGCCTTGCATTCTTGTAGCCATGTGCTGCAT
GATCAATGCTTGCCGTTCTTGCA

>Hs_sRNA_23.2 Hs_sRNA_23.2 undefined product 747547:747651 reverse
GCATCGGCTTTTTTATTGCCCGCTTGCTTGCATCATGGAGCCATCGGGCCGAGGCCG
CACAATGGCTCGCATTGCGCGTTTTGCTGCCCTGCAAACTTTTT

APÊNDICE II – SCRIPT PARA *TRIMMING* DE LEITURAS DE RNA-SEQ

```
#!/usr/bin/perl -w
# Escrito por Vinicius Weiss – UFPR - 2012

$file = shift or die "Usage: meu.pl <fasta>\n";
$out=shift or die "arquivo de saida\n";
$seed=shift or die "\n especifique quanto deseja cortar";
open (AMOSTRA , "<$file");
open OUT, (">$out");
%reads;
while ($line=<AMOSTRA>){
    chomp $line;
    if ($line=~/#/) {

        #print ARQ "$line\n";
    }elseif ($line=~/\>/){
        $id=$line;
    }else{
        $reads{$id}=$line;
    }
}
print "ETAPA 1 --OK\n";
foreach $key (sort keys %reads) {
    $seq= $reads{$key};
    $tam=length($seq);
    ($seq)=(substr($seq,0,($tam-$seed+1)));
    print OUT "$key\n$seq\n";
}
}
```


APÊNDICE III – ALVOS PREDITOS ATRAVÉS DA FERRAMENTA TARGETRNA

NCRNA	ALVOS PREDITOS						
Hs_noco_471	mltB	Hsero_2631	modC1	coaE	Hsero_4755	Hsero_2152	Hsero_0520
Hs_noco_494	secA	Hsero_4143	Hsero_0875	Hsero_0887	Hsero_1995		
Hs_noco_525	etf	adhC	argG	Hsero_1800			
Hs_noco_530	Hsero_1355	Hsero_0259					
Hs_noco_534	umuC	Hsero_2044	Hsero_2895	potC	pilW	Hsero_2930	Hsero_0515
Hs_noco_637	Hsero_1179	pbpC	Hsero_3618	Hsero_0075	dgoK	aidB	Hsero_1443
Hs_noco_651	Hsero_2659	benE	Hsero_3225	Hsero_1374	Hsero_2212		
Hs_noco_682	Hsero_2384	antB	Hsero_2745	Hsero_4732	abgB	Hsero_4708	Hsero_4711
Hs_noco_698	Hsero_2805	Hsero_0916	miaB	Hsero_0226	nahD	Hsero_2484	Hsero_3718
Hs_noco_700	Hsero_0937	Hsero_3351	Hsero_3273	Hsero_3431			
Hs_noco_709	cobH	Hsero_3539	Hsero_4243	Hsero_2064	alr	Hsero_0619	Hsero_2490
Hs_noco_719	Hsero_2631	Hsero_4466	mutT	nifR3	Hsero_0386	poxR	rimL
Hs_noco_722	Hsero_1742	Hsero_1962	Hsero_0326	crp	Hsero_0021		
Hs_noco_723	glcD	degQ	Hsero_0237	ndeD			
Hs_noco_726	Hsero_1276	Hsero_4201	selD	rhtB	Hsero_4012		
Hs_noco_732	Hsero_3615	Hsero_0190	fecR	recR	coxB	Hsero_0521	paaN
Hs_noco_742	Hsero_4078	nifN	ilvA	Hsero_0831	Hsero_4670	Hsero_2387	Hsero_3033
Hs_noco_744	fdx	Hsero_1868	Hsero_2366	Hsero_1224	Hsero_3711	vgr	epsB
Hs_noco_746	hydH	vgr	Hsero_2281	qor	Hsero_2152	Hsero_1881	Hsero_3847
Hs_noco_748	Hsero_2276	Hsero_2371	Hsero_3234	Hsero_3819	Hsero_1150	livM	Hsero_1502
Hs_noco_749	Hsero_0347	Hsero_0248	hesB1	Hsero_4273	Hsero_4230	pstB	Hsero_0169
Hs_noco_750	vsrB	tdcF	Hsero_2622	fabG	Hsero_4266		
Hs_noco_761	Hsero_4684	pcm	Hsero_0582	Hsero_3041	Hsero_3656	Hsero_1268	adh
Hs_noco_802	Hsero_4705	Hsero_0458	Hsero_2877	yedZ	Hsero_1154	Hsero_2955	Hsero_0040
Hs_noco_820	fabG	Hsero_1811	Hsero_0252	Hsero_0386			
Hs_sRNA_4.2	Hsero_3259	Hsero_2950					
Hs_sRNA_19.2	Hsero_1635	Hsero_0230	katG	Hsero_2975	Hsero_1603	livG	Hsero_2400
Hs_sRNA_2.1	Hsero_2659	Hsero_4398	Hsero_3780	benE	Hsero_0619	Hsero_4622	murF
	Hsero_0730	Hsero_4518	glcE	rpoE	Hsero_3746	narX	
Hs_sRNA_Gs10	psd						
Hs_sRNA_Gs12	rpoE	murB	Hsero_1865	Hsero_4625	Hsero_0803	Hsero_3627	Hsero_2196
Hs_sRNA_Gs36	Hsero_0361	fdx	Hsero_1175	catB2	dtd	Hsero_3198	
Hs_noco_811	Hsero_4012	Hsero_3787	Hsero_4645	Hsero_3447	mdoG	Hsero_1834	Hsero_4261

Hs_noco_498	paaD	Hsero_3188	Hsero_0567	rpsE	Hsero_0781	Hsero_1581	spy
Hs_noco_780	ogT	livF	Hsero_3651	Hsero_0665	Hsero_4708	tctE	
Hs_noco_786	glgA	nasA	Hsero_0788	Hsero_3670	Hsero_4681		
Hs_noco_757	Hsero_0361	fdx	narL	ssb	Hsero_1175		
Hs_noco_725	Hsero_3979	cspD	purK	fleN	rpsN	Hsero_3525	Hsero_3570
Hs_noco_714	Hsero_2906	mqo	mreC	Hsero_3947			
Hs_noco_640	ubiD	fdxA	Hsero_1672	Hsero_2671	petC	ahcY	Hsero_0544
Hs_noco_602	uvrD	Hsero_2485	Hsero_2004	Hsero_4374	alkB	araJ	livH
Hs_noco_564	tonB	Hsero_3653	livF	fnr	Hsero_1086	Hsero_2271	Hsero_2660
Hs_noco_565	Hsero_4424	Hsero_1893	Hsero_3062				
Hs_noco_568	Hsero_4024	Hsero_2373	rimM	Hsero_4056	Hsero_2156	Hsero_2755	
Hs_noco_593	Hsero_2208	cmk					
Hs_noco_551	gcpE	Hsero_2937	rhIE	Hsero_0632	Hsero_3323		
Hs_noco_555	Hsero_2659	gltP	benE	Hsero_3148	Hsero_3412	Hsero_1499	Hsero_2621
Hs_noco_526	obgE	Hsero_1799					
Hs_noco_490	Hsero_2301	Hsero_0648	Hsero_1062	ompW2	fhaC	Hsero_3527	glnP
Hs_noco_514	dppA	Hsero_2711	rpsJ	Hsero_2351	Hsero_3461	Hsero_4794	Hsero_1729
Hs_noco_724	Hsero_4090	pyrD	Hsero_0598	glpR	degQ	htpX	
"CUFF.55"	rsuA	Hsero_2805	Hsero_2877				
"CUFF.111"	ygiY	Hsero_0269	nasD	Hsero_2007	Hsero_2687	Hsero_3572	
"CUFF.124"	Hsero_3672	Hsero_0238	Hsero_4250	Hsero_2521	livK	uppS	
"CUFF.223"	Hsero_2659	betA	cheW	Hsero_4428			
"CUFF.251"	Hsero_4773	trxB2	Hsero_4170	Hsero_4503			
"CUFF.273"	tauD	Hsero_1839	Hsero_1637	Hsero_3317	araJ	aer	Hsero_2700
"CUFF.395"	Hsero_3726	dadA	Hsero_4471	dctQ			
"CUFF.404"	Hsero_4775	Hsero_4087	entB	Hsero_2757	gidB	phaZ	plcN
"CUFF.504"	Hsero_4775	Hsero_4087					
"CUFF.522"	Hsero_3899	Hsero_1245	Hsero_0645	livF			
"CUFF.524"	Hsero_1148	Hsero_2801	Hsero_2106	livG	Hsero_3871	caiB	Hsero_1949
"CUFF.689"	gcp	Hsero_0385	tauB	uup	Hsero_1722		
"CUFF.690"	Hsero_3780	Hsero_1416	Hsero_1693	Hsero_1374	Hsero_4018	Hsero_4074	
"CUFF.738"							
"CUFF.829"							
"CUFF.852"	Hsero_3438	dctQM	wrbA	Hsero_1176	Hsero_4712	Hsero_2376	Hsero_1726
"CUFF.934"	Hsero_2699	Hsero_0021	Hsero_3356	Hsero_3059	fdsG	Hsero_2738	Hsero_3347
"CUFF.938"	Hsero_3699	Hsero_4574	Hsero_2326	bioA	Hsero_3190	Hsero_3904	Hsero_3361
"CUFF.967"	prmA	Hsero_3145	Hsero_1148	Hsero_1370	cheZ	creA	citE

"CUFF.1011"							
"CUFF.1094"	Hsero_4708	rimL	rfbB	Hsero_4764	Hsero_2890	spoT	tctC
"CUFF.1117"	narH	tyrB	Hsero_0885	ssuC	papD	gdhA	degQ
"CUFF.1142"	Hsero_4323	Hsero_2346					
"CUFF.1223"	Hsero_3628	Hsero_1467	Hsero_3308	mrcA	spoT	pilB	Hsero_4248
"CUFF.1716"	phnX	recX	Hsero_3656	gcpE			
"CUFF.598"	leuC	thrB	Hsero_3039	ompR	tdcB	Hsero_3668	cdsA
"CUFF.225"	Hsero_4032	Hsero_1076	Hsero_1531	Hsero_0899	Hsero_3440	Hsero_2358	glcE
"CUFF.228"	Hsero_0216	Hsero_1062	Hsero_1146	cyoB	Hsero_3074	glgB	Hsero_2267
"CUFF.1027"	Hsero_4588	Hsero_3837	uvrC	mrdA	Hsero_1499	Hsero_1811	flhF
"CUFF.1106"	Hsero_3619	fiu	Hsero_3751	Hsero_4802	pilB	Hsero_3285	Hsero_1206
"CUFF.1109"							
"CUFF.1158"	aro8	Hsero_3302	Hsero_3511	Hsero_4740	pilN		
"CUFF.1190"	Hsero_2589	Hsero_3606	hutH	recR	uup	Hsero_0743	atoC
"CUFF.1194"	urtB	Hsero_0779	Hsero_1695	Hsero_1623	hemK		
"CUFF.1510"	Hsero_3048	Hsero_3392	Hsero_4174	dksA			
"CUFF.1548"	dacC	Hsero_2750	cheR	Hsero_2198	ompC		
"CUFF.1570"							
CUFF.1692	fdx	ackA	fold	Hsero_1013	Hsero_2521		
"CUFF.1713"	htrB	recG					
"CUFF.1720"	htrB	Hsero_4748	Hsero_1571	Hsero_0618			
"CUFF.1751"	kdtA	dadA	speE	Hsero_4467	Hsero_0393	Hsero_1445	Hsero_1640
Hs_sRNA_23.2	Hsero_2721	maf	Hsero_4137	Hsero_2372	gspG	Hsero_2340	Hsero_2496
Hs_noco_819	Hsero_4684	pcm	Hsero_0582	Hsero_3041	Hsero_3656	Hsero_1268	adh
Hs_noco_656	Hsero_2805	Hsero_0916	miaB	Hsero_0226	nahD	Hsero_2484	Hsero_3718
Hs_noco_764	Hsero_0222	serA	Hsero_0901	Hsero_1683	parA		
Hs_noco_652	Hsero_2384	antB	Hsero_2745	Hsero_4732	abgB	Hsero_4708	

APÊNDICE IV – SCRIPT EM BASH PARA EXECUTAR A FERRAMENTA CMSERACH

```
#!/bin/bash
for rfam in `ls files/sequences_ncRNA/sequences`
do
#echo $rfam
./tools/infernal-1.0.2/src/cmsearch -T 30 files/sequences_ncRNA/sequences/${rfam} > files/rfam/ncRNA/${rfam}.out
echo processando...
done
exit 0
runinfernal.sh (END)
```