UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

JOÃO PAULO PEREIRA DE ALMEIDA

O transcrito antisense primário de *Halobacterium salinarum* NRC-1

Ribeirão Preto
2018

JOÃO PAULO PEREIRA DE ALMEIDA

O transcrito antisense primário de *Halobacterium salinarum* NRC-1

Versão Original

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Mestre em Ciências.

Área de Concentração: Bioquímica

Orientadora: Profª. Dra. Tie Koide

Coorientador: Prof. Dr. Ricardo Zorzetto Nicoliello Vêncio

Ribeirão Preto
2018

Autorizo, a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na publicação
Serviço de Biblioteca e Documentação
Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

Almeida, João Paulo Pereira

O transcrito antisense primário de *Halobacterium salinarum* NRC-1.

João Paulo Pereira de Almeida; orientadora, Tie Koide. Ribeirão Preto - 2018

127 f. :il. 30cm

Dissertação (Mestrado em Ciências) – Programa de Pós-Graduação em Bioquímica, Universidade de São Paulo, Ribeirão Preto, 2018.

Versão Original

1-Domínio Archaea; 2-*Halobacterium salinarum* NRC-1; 3-RNAs não-codificantes; 4-RNAs antisense

Nome: ALMEIDA, João Paulo Pereira

Título: O transcrito antisense primário de *Halobacterium salinarum* NRC-1

Dissertação apresentada ao Programa de Pós-Graduação em Bioquímica da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo, para obtenção do Título de Mestre em Ciências.

Aprovado em: ____/____/____

Banca Examinadora

Prof(a). Dr (a). _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Prof(a).Dr (a). _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Prof(a). Dr (a). _____

Instituição: _____

Julgamento: _____

Assinatura: _____

Apoio e Suporte Financeiro

Este projeto foi realizado com o apoio financeiro das seguintes entidades:

Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo: 2015/21038-1)

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES

Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq

Fundação de Apoio ao Ensino, Pesquisa e Assistência – FAEPA

Dedico esse trabalho os meus pais, Paulo e Vilma pela confiança e apoio incondicional aos meus estudos.

Agradecimentos

Agradeço a minha família, meus pais Vilma e Paulo e minha irmã Mariana, pelo apoio e conforto que sempre me deram ao longo de meus estudos.

Agradeço a minha namorada Gabriella, pelos 6 anos de companheirismo e por compartilhar as tristezas e alegrias dessa jornada.

Agradeço a Prof^a Tie Koide e ao Prof. Ricardo Vêncio, pela confiança e liberdade que me deram para trabalhar durante o mestrado. O rigor científico com que trabalham é algo que levarei comigo para sempre.

Agradeço a Sílvia Helena Epifânio, por toda dedicação e carinho com que se dedica ao laboratório e aos seus integrantes. Você é um exemplo de funcionário público e de pessoa que realmente faz a diferença onde trabalha.

Agradeço aos colegas de labisismi da “geração anterior”, Caten e Vicente. Foi um prazer enorme conviver com vocês.

Agradeço aos colegas de labisismi da “nova geração”, Alan, Junier, Vinícius e Catarina. Vocês fazem do labisismi um ambiente extremamente agradável e rico em discussões científicas e críticas. Tem sido um grande prazer trabalhar ao lado de vocês.

Agradeço ao Luander. Pelas ótimas conversas e aprendizado que o projeto paralelo em astrobiologia me proporcionou.

Agradeço aos colegas do labpib, Yagoub, Amr e Danillo. Por todas discussões, conversas e apoio técnico.

Agradeço a todos colegas da bioquímica com os quais tive o prazer de conviver durante as disciplinas, coordenação de curso de inverno e intervalos para um café, isso inclui os “vizinhos de corredor” integrantes dos laboratórios do Prof. Adriano, Prof. Marcelo, Prof. Cláudio, Prof. Roberto, Prof. Vitor e Prof. Eduardo, que sempre estiveram de prontidão para prestar um auxílio técnico, discutir experimentos e emprestar reagentes. Em especial aos colegas André, Sarah, Diego, Nichelle, Dimitrius, Mariana, Tatiana e Matheus.

Agradeço a Maria Ivone, secretária da pós-graduação e aos secretários do departamento Ronaldo e Rose. Pela paciência e prontidão nos momentos em que precisei de ajuda.

Resumo

ALMEIDA, João Paulo Pereira. **O transcrito antisense primário de *Halobacterium salinarum* NRC-1**. 2018, 127 f. Dissertação (Mestrado em Bioquímica) – Faculdade de Medicina de Ribeirão Preto – Universidade de São Paulo, Ribeirão Preto, 2018.

Em procariotos, RNAs antisense (asRNAs) constituem a classe de RNAs não codificantes (ncRNAs) mais numerosa detectada por métodos de avaliação de transcrito em larga escala. Apesar da grande abundância, pouco se sabe sobre mecanismos regulatórios e aspectos da conservação evolutiva dessas moléculas, principalmente em arqueias, onde o mecanismo de degradação de RNAs dupla fita (dsRNAs) é um fenômeno pouco conhecido. No presente estudo, utilizando dados de dRNA-seq, identificamos 1626 inícios de transcrição primários antisense (aTSSs) no genoma de *Halobacterium salinarum* NRC-1, importante organismo modelo para estudos de regulação gênica no domínio *Archaea*. Integrando dados de expressão gênica obtidos a partir de 18 bibliotecas de RNA-seq *paired-end*, anotamos 846 asRNAs a partir dos aTSSs mapeados. Encontramos asRNAs em ~21% dos genes anotados, alguns desses relacionados a importantes características desse organismo como: codificadores de proteínas que constituem vesículas de gás e da proteína bacteriorodopsina, além de vários genes relacionados a maquinaria de tradução e transposases. Além desses, encontramos asRNAs em genes pertencentes a sistemas de toxinas-antitoxinas do tipo II e utilizando dados públicos de dRNA-seq, evidenciamos que esse é um fenômeno que ocorre em bactérias e arqueias. A interação de um ncRNA com seu RNA alvo pode ser dependente de proteínas, em arqueias, a proteína LSm é uma chaperona de RNA homóloga a Hfq de bactérias, implicada no controle pós-transcricional. Utilizamos dados de RIP-seq de RNAs imunoprecipitados com LSm e identificamos 91 asRNAs interagindo com essa proteína, para 81 desses, o mRNA do gene sense também foi encontrado interagindo. Buscando por aTSSs presentes nas mesmas regiões de genes ortólogos, identificamos 160 aTSSs que dão origem a asRNAs em *H. salinarum* possivelmente conservados em *Haloferax volcanii*. A expressão dos asRNAs anotados foi avaliada ao longo de uma curva de crescimento e em uma linhagem *knockout* de um gene que codifica uma RNase R, possível degradadora de dsRNAs em arqueias. Encontramos um total de 144 asRNAs diferencialmente expressos ao longo da curva de crescimento, para 56 desses o gene sense também está diferencialmente expresso, caracterizando possíveis mecanismos de regulação em *cis* por esses RNAs. Na linhagem *knockout*, encontramos cinco asRNAs diferencialmente expressos e apenas para um desses o gene sense também está diferencialmente expresso, resultado que não nos permitiu inferir um possível papel de degradação de dsRNAs da RNase R em *H. salinarum* NRC-1. Nesse trabalho apresentamos um mapeamento completo do transcrito antisense primário de *H. salinarum* NRC-1 com resultados que consistem em um importante passo na direção da compreensão do envolvimento da transcrição antisense na regulação gênica pós-transcricional desse organismo modelo do terceiro domínio da vida.

Palavras-chave: RNAs antisense (asRNAs). RNAs não codificantes (ncRNAs). Differential RNA-seq (dRNA-seq). Inícios de transcrição primários (TSSs). RNAs dupla fita (dsRNAs). RNase R. Sistemas toxinas-antitoxinas (TAs). Lsm. *Halobacterium salinarum*. *Archaea*.

Abstract

ALMEIDA, João Paulo Pereira. **The antisense primary transcriptome of *Halobacterium salinarum* NRC-1**. 2018, 127 f. Dissertation (Master in Biochemistry) – Ribeirão Preto Medical School – University of São Paulo, Ribeirão Preto, 2018.

Antisense RNAs (asRNAs) constitute the most numerous class of non-coding RNAs (ncRNAs) detected by transcriptome highthroughput methods in prokaryotes. Despite this abundance, little is known about regulatory mechanisms and evolutionary aspects of these molecules, mainly in archaea, where the mechanism of double-strand RNA (dsRNA) degradation remains poorly understood. In this study, using dRNA-seq data, we identified 1626 antisense transcription start sites (aTSSs) in the genome of *Halobacterium salinarum* NRC-1, an important model organism for gene expression regulation studies in *Archaea*. By integrating gene expression data from 18 RNA-seq paired-end libraries, we were able to annotate 846 asRNAs from mapped aTSSs. We found asRNAs in ~21% of annotated genes including genes related to important characteristics of this organism, such as: gas vesicle proteins, bacteriorhodopsin, translation machinery and transposases. We also found asRNAs in type II toxin-antitoxin systems and using public dRNA-seq data, we show evidences that this phenomenon might be conserved in archaea and bacteria. The interaction of a ncRNA with its target may depend on intermediary proteins action. In archaea, the LSm protein is a RNA chaperone homologous to bacterial Hfq, involved in post-transcriptional regulation. We used RIP-seq data from RNAs immunoprecipitated with LSm and identified 91 asRNAs interacting with this protein, for 81 of these the mRNA of the sense gene is also interacting. We searched for aTSSs present in the same region of orthologous genes in the *Haloferax volcanii*. We found 160 aTSSs that originated asRNAs in *H. salinarum* NRC-1 that might be conserved in this two archaea. The expression of annotated asRNAs was analyzed over a growth curve and in a *knockout* strain for RNase R gene. We found 144 asRNA differentially expressed over the growth curve, for 56 of these the sense gene was also differentially expressed, characterizing possible *cis* regulators asRNAs. In the knockout strain we found five differentially expressed asRNAs and only one asRNA/gene pair, this result does not allow us to infer a dsRNA degradation *in vivo* activity for this RNase in *H. salinarum* NRC-1. This work contributes to the discovery of the antisense transcriptome in *H. salinarum* NRC-1 a relevant step to uncover the post-transcriptional gene regulatory network in this archaeon.

Keywords: Antisense RNAs (asRNAs). Non-coding RNAs (ncRNAs). Differential RNA-seq (dRNA-seq). Transcription start sites (TSSs). Double-strand RNAs (dsRNAs). RNase R. Toxins-antitoxins systems (TAs). LSm. *Halobacterium salinarum*. Archaea.

Lista de ilustrações

Figura 1- Os três domínios da vida.....	19
Figura 2- <i>H. salinarum</i>	20
Figura 3- Bacteriorodopsina.....	21
Figura 4- Bomba de prótons.....	22
Figura 5- Mutantes de vesículas de gás.....	24
Figura 6- <i>vac regions</i>	24
Figura 7- Possíveis modos de ação de Hfq.....	32
Figura 8- Transcritos sobrepostos em procariotos.....	33
Figura 9- dRNA-seq.....	36
Figura 10- Efeitos co-transcricionais da transcrição antisense.....	38
Figura 11- Regulação pós-transcricional por asRNAs.....	39
Figura 12- Sistemas Toxina-antitoxina.....	42
Figura 13- lasRNAs.....	43
Figura 14- Correlação inversa dos níveis de transcritos sense e antisense em <i>H. volcanii</i>	44
Figura 15- Distribuição de valores de MFE calculados para sequências de 51nt extraídas a cada 10nt (janela deslizante) ao longo do genoma de <i>H. salinarum</i> NRC-1.....	50
Figura 16- Pipeline usado para anotação dos asRNAs.....	51
Figura 17- Posições relativas dos aTSSs.....	59
Figura 18- <i>Screenshot</i> GGB.....	60
Figura 19- Dados de bibliotecas PE para anotação de asRNAs.....	62
Figura 20- asRNA sobrepondo dois genes.....	63
Figura 21- Distribuição de tamanho dos 846 asRNAs loci definidos a partir dos aTSSs.....	64
Figura 22- Frequência nucleotídica das extremidades 3' dos asRNAs anotados.....	64
Figura 23- Distribuição da distância em nt da extremidade 3' dos asRNAs anotados até a extremidade 5' do gene cognato.....	65
Figura 24- Cobertura de reads dos asRNAs em relação a seus genes cognatos.....	66
Figura 25- Análise da frequência nucleotídica nas regiões promotoras e de início de transcrição dos asRNAs anotados.....	67
Figura 26- asRNAs loci anotados para os genes VNG_RS00195 e VNG_RS13220.....	68
Figura 27- Representatividade dos KOs identificados nos genes com asRNAs anotados.....	69
Figura 28- asRNA anotado para o gene brb.....	72
Figura 29- asRNAs presentes nos genes <i>gvpD</i> , <i>gvpA</i> e <i>gvpC</i> da <i>vac region</i> <i>gvp1</i>	74

Figura 30- asRNA e aTSS identificados para o gene yoeB que codificada uma toxina em um TAs do tipo II em <i>E. coli</i>	75
Figura 31- asRNAs anotados em um sistema TAs do tipo II em <i>H. salinarum</i> NRC-1.....	76
Figura 32- Pontos de amostragem na curva de crescimento.....	77
Figura 33- Pares asRNAs/genes diferencialmente expressos na passagem de 17H para 37H..	78
Figura 34- Pares asRNAs/genes diferencialmente expressos na passagem de 37H para 86H..	79
Figura 35- 56 pares asRNA/gene diferencialmente expressos em ao menos uma das duas transições da curva de crescimentos analisada.....	80
Figura 36- Exemplos de aTSSs conservados em <i>H. salinarum</i> NRC-1 e <i>H. volcanii</i>	82
Figura 37- Resultado da busca por domínios proteicos nas RNases R utilizando BlastCD.....	83
Figura 38- Alinhamento múltiplo das sequências de aa das RNases R de <i>H. salinarum</i> NRC-1 e <i>H. volcanii</i>	84
Figura 39- Curva de crescimento.....	84
Figura 40- Par asRNA/gene upregulado na linhagem knockout para o gene VNG2647G.....	85
Figura 41- 3'UTR antisense longo.....	91

Lista de tabelas

Tabela 1 – Número de reads sequenciados, trimados e alinhados para cada biblioteca.....	61
Tabela 2 – asRNAs em genes relacionados a tradução.....	70
Tabela 3 – asRNAs sobrepostos a extremidade 5' de genes codificadores de transposases.....	71
Tabela 4 – asRNAs nos genes codificadores e reguladores de rodopsinas.....	73
Tabela 5 – asRNAs nos genes gvp da região vac gvp-1.....	74
Tabela 6: Número de TAs do tipo II previamente anotados e anotados nesse estudo com a ferramenta TA finder.....	75
Tabela 7: TAs do tipo II com aTSSs mapeados.....	76
Tabela 8: asRNAs diferencialmente expressos na linhagem <i>knockout</i> para o gene VNG2647G.	85

Lista de Siglas

asRNA - *antisense RNA*; RNA antisenso

ATP - *adenosine triphosphate*; trifosfato de adenosina

aTSS - *antisense transcript start site*; início de transcrição primário antisense

cDNA - *complementary DNA* – DNA complementar

CDS - *coding sequence*; sequência codificadora

CM - *complex media*; meio complexo

CRISPR - *clustered regularly interspaced short palindromic repeats*; agrupados de curtas repetições palindrômicas regularmente interespçadas

crRNA - *Clustered Regularly Interspaced Short Palindromic Repeats RNAs*

daTSS - *downstream TSS*

DNA - *deoxyribonucleic acid*; ácido desoxirribonucleico

dRNA-seq - *differential RNA-seq*; RNA-seq diferencial

dsRNA - *double strand RNA* – RNA dupla fita

FC - *fold change*

GFF - *general feature file*

GGB - *Gaggle genome browser*

IS - *insertion sequence*; sequência de inserção

lasRNA - *long non coding asRNA*; asRNA não-codificante longo

LB - Luria bertani

lncRNA - *long non coding RNA*; RNA não-codificante longo

MFE - *minimum free energy*; mínima energia livre

miRNA - *micro RNA*

mRNA - *messenger RNA*, RNA mensageiro

NCBI - *National Center for Biotechnology Information*

ncRNA - *non coding RNA*; RNA não-codificante

nt - *nucleotídeos*

OD - *optical density*; densidade óptica

ORF - *Open Reading Frame*; janela aberta de leitura

oTSS - *orfán transcript start site* – início de transcrição primário órfão

PCR - *polymerase chain reaction* – Reação em cadeia da polimerase

PE - *paired-end*

pTSS - *primary transcript start site* – início de transcrição primário

RBP - *RNA binding protein*; proteína ligante a RNA

RBS - *ribosomal binding site*; sítio de ligação ribossomal

RIP-seq - *RNA co-immunoprecipitation sequencing*; sequenciamento de RNAs co-imunoprecipitados

RIP-chip - *RNA co-immunoprecipitation microarray*; microarray de RNAs co-imunoprecipitados RNA - *ribonucleic acid*; ácido ribonucleico

RNAi - interference RNA; RNA de interferência

RNA-seq - *RNA sequencing*; sequenciamento de RNA

rRNA - *ribosomal RNA*; RNA ribossomal

SE - *single-end*

SD - Shine Dalgarno

siRNA - *small interfering RNA*; pequeno RNA de interferência

snoRNA - *small nucleolar RNA*; pequeno RNA nucleolar

snoRNP - *small nucleolar ribonucleoprotein*; Ribonucleoproteína de ligação a pequenos RNAs nucleolares

sotRNA - *sense overlapping transcripts RNAs*; RNAs sobrepostos a transcritos sense.

SRA - *Sequence Read Archive*

sRNA - *small RNA*; pequeno RNA

SRP - *signal recognition particle*; partícula de reconhecimento de sinal

TAs - *Toxin-antitoxin system*; sistema toxina-antitoxina

TER - *Telomerase RNA*

TEX - *Terminator 5'-Phosphate-Dependent Exonuclease*

tmRNA - *transfer-messenger RNA*

tRNA – *transfer RNA*; RNA transportador

TSS - *transcription start site*; sítio de início da transcrição

TSSaRNA - *transcription start site associated RNA*; RNA associado ao sítio de início da transcrição

UTR - *untranslated region*; região não traduzida

Sumário

1-Introdução.....	17
1.1-Domínio Archaea.....	17
1.1.1- <i>Halobacterium salinarum</i> NRC-1.....	19
1.2-RNAs não-codificantes.....	26
1.3-RNAs antisense.....	36
2-Objetivo Geral.....	46
2.1-Objetivos específicos.....	46
3-Materiais e métodos.....	47
3.1-dRNA-seq.....	47
3.1.1-Linhagem e condição de cultivo.....	47
3.1.2-Extração RNA dRNA-seq.....	47
3.1.3-Depleção de RNAs processados.....	47
3.1.4-Preparação das bibliotecas de dRNA-seq.....	48
3.2-Processamento dos dados de RNA-seq.....	48
3.3-Mapeamento de TSS em <i>H. salinarum</i> NRC-1.....	48
3.4-Filtragem de aTSS em <i>H. salinarum</i> NRC-1.....	49
3.5-Anotação dos loci de asRNAs.....	50
3.6-Avaliação de sequências consenso na região promotora e extremidade 3'.....	52
3.7-RIP-seq.....	52
3.7.1-RIP-seq (RNA Immunoprecipitation – RNA-seq).....	52
3.7.2-RNA-seq de RNAs co-imunoprecipitados com a proteína LSm (RIP-Seq).....	53
3.7.3-Análise RIP-seq.....	53
3.8-asRNAs interagindo com LSm.....	54
3.9-Função dos genes sense.....	54
3.10-Anotação de TAs tipo II e busca de aTSSs.....	54
3.11-Expressão diferencial.....	55
3.12-aTSSs conservados em <i>H. volcanii</i>	55
3.13-Comparação entre sequências de RNases R.....	56
3.14 Construção da linhagem Δ naseR.....	56
3.15-Curva de crescimento.....	56
3.16-Extração RNA total.....	57
3.17-RNA-seq Δ naseR.....	57
4-Resultados.....	59
4.1-Mapeamento da extremidade 5' de asRNAs: aTSSs.....	59
4.2-Anotação dos asRNAs: inferência das extremidades 3'.....	60
4.3-asRNAs sobrepostos a extremidade 5' dos genes.....	64
4.4-Expressão em relação ao gene sense.....	65
4.5-Região Promotora.....	67
4.6-asRNAs que interagem com LSm.....	67
4.7-Função dos genes sense potencialmente regulados por asRNAs.....	68
4.7.1-asRNAs em genes relacionados a Tradução.....	69
4.7.2-asRNAs em genes codificadores de Transposases.....	69

4.7.3-asRNAs nos genes codificadores de Rodopsinas.....	72
4.7.4-asRNAs em genes relacionados a vesículas de gás – (<i>gas vescicle proteins</i> - gvp).....	73
4.7.5-asRNAs em TAs tipo II em bactérias e arqueias.....	74
4.8-Expressão diferencial na curva de crescimento.....	77
4.9-aTSSs conservados em <i>H. volcanii</i>	81
4.10-RNase R.....	82
5-Discussão.....	86
5.1-Mapeamento de aTSSs e anotação de asRNAs: mapeamento das extremidades 5' e inferência das extremidades 3'.....	86
5.2- asRNAs sobrepostos a extremidade 5' dos genes.....	91
5.3-Expressão em relação ao gene sense.....	93
5.4-Região Promotora.....	94
5.5- asRNAs que interagem com LSm.....	95
5.6-Função dos genes sense potencialmente regulados por asRNAs.....	97
5.6.1-Tradução.....	97
5.6.2-Transposases.....	98
5.6.3-asRNAs nos genes codificadores de Rodopsinas.....	98
5.6.4-asRNAs em genes relacionados a vesículas de gás – (<i>gas vescicle proteins</i> – gvp).....	99
5.7-asRNAs em TAs tipo II em bactérias e arqueias.....	101
5.8-Expressão diferencial na curva de crescimento.....	102
5.9-aTSSs conservados em <i>H. volcanii</i>	104
5.10-RNase R.....	106
6-Conclusões.....	108
7-Referências.....	109
Apêndice.....	126
Apêndice A – Materiais suplementares.....	126

1- Introdução

1.1 - Domínio Archaea

Todos organismos vivos pertencem a um dos três grupos, *Eukarya*, *Bacteria* ou *Archaea*, chamados de domínios da vida, esses representam os três grandes ramos da árvore filogenética da vida em nosso planeta (Figura 1). O domínio Archaea, também chamado de “terceiro domínio da vida”, foi o último a ser proposto. Em 1977, analisando sequências nucleotídicas da subunidade menor do RNA ribossomal para estabelecer relações filogenéticas, Woese e Fox dividiram os procariotos em dois grandes domínios, dando origem ao domínio Archaeobacteria que viria a ser renomeado para Archaea (da palavra grega ἀρχαῖα, “antigo”) (WOESE; FOX, 1977; WOESE; KANDLER; WHEELIS, 1990). Provavelmente, arqueias e bactérias divergiram cedo no curso evolutivo da vida em nosso planeta e os eucariotos evoluíram a partir do mesmo ramo que deu origem as arqueias, sendo que, muitos estudos bioquímicos e genéticos suportam a maior proximidade filogenética de arqueias de eucariotos do que de bactérias (CHABAN; NG; JARRELL, 2006). Uma das hipóteses para o surgimento dos eucariotos é a de que seus ancestrais se originaram quando uma relação endossimbiótica foi estabelecida entre células de arqueias e alfa-proteobactérias (grupo de bactérias do qual se originaram as mitocôndrias) (MARTIN; GARG; ZIMORSKI, 2015). Além disso, esse domínio abriga muitos organismos capazes de viver em condições extremas, similares as condições primitivas do planeta Terra durante o éon Arqueano, o que leva a hipótese de estarem estreitamente relacionados a origem da vida em nosso planeta (CAVICCHIOLI, 2011).

O domínio Archaea é dividido em quatro filos principais (Figura 1): Euryarchaeota, o mais diverso, composto por arqueias metanogênicas, halófilas, termófilas e psicrófilas; Crenarchaeota, representado por organismos hipertermófilos, com as espécies capazes de sobreviver as condições extremas de temperatura; Nanoarchaeota e Korarchaeota com poucos representantes e posicionamento filogenético ainda incerto (ALLERS; MEVARECH, 2005)

Arqueias apresentam um mosaico de características compartilhadas com eucariotos e bactérias. Muitas de suas vias metabólicas e seus genomas compactos com vários genes dispostos em *operons* são características que compartilham com bactérias. Quanto aos processos de replicação, transcrição e tradução, apresentam genes similares aos de eucariotos

(ALLERS; MEVARECH, 2005; GRABOWSKI; KELMAN, 2003). Integrantes desse domínio também apresentam características exclusivas como a composição da bicamada lipídica. A cadeia lateral de hidrocarbonetos derivados de ácidos graxos é unida por ligações éster a um grupo glicerol-3-fosfato na membrana plasmática de bactérias e eucariotos. Na membrana plasmática de arqueias, ocorrem cadeias laterais de isoprenóides unidas por ligações éter a um grupo glicerol-1-fosfato (JAIN; CAFORIO; DRIESSEN, 2014). Essa composição afeta a estabilidade da membrana plasmática desses organismos, sendo relacionada a capacidade de adaptação a condições ambientais hostis, como a de altas temperaturas. Diferentemente das bactérias, não possuem peptidoglicanos em suas paredes celulares. Arqueias também não são capazes de formar esporos ou realizar fotossíntese (MOISSEL-EICHINGER et al., 2018). Além dessa característica, Archaea é o único domínio em que se encontra organismos metanogênicos, o que lhes garante papel ecológico fundamental em ciclos biogeoquímicos e como simbiontes (CAVICCHIOLI, 2011; MOISSEL-EICHINGER et al., 2018).

Figura 1- Os três domínios da vida. A divisão da vida nesse planeta em três domínios baseada nas sequências da subunidade menor do rRNA foi proposta por Carl Woese em 1977. O domínio Archaea é dividido em quatro filos principais: Euryarchaeota, Crenarchaeota, Nanoarchaeota e Korarchaeota (Fonte: (ALLERS; MEVARECH, 2005))

Algumas das adaptações moleculares de archaeas a ambientes extremos são fontes de matéria prima para aplicações biotecnológicas. A DNA polimerase Pfu, extraída da arqueia hipertemófila *Pyrococcus furiosus*, mantém sua atividade catalítica em altas temperaturas, sendo utilizada rotineiramente em laboratórios para reações de PCR (BILES; CONNOLLY, 2004). Ambientes de conversão de biomassa possuem altas concentrações iônicas e são inibitórios para muitas enzimas. As enzimas de arqueias halófilas são adaptadas a essa condição e são estudadas como possíveis alternativas (ZHANG, TAO et al., 2011).

Por muito tempo, arqueias receberam a alcunha de organismos pertencentes a ambientes extremos, porém com o emprego das novas tecnologias de sequenciamento que possibilitaram a detecção de genomas completos de organismos não cultiváveis, essa perspectiva mudou e a ubiquidade desses organismos se tornou evidente. Além de questões voltadas para a origem da vida, biotecnologia e ecologia, a influência desses organismos na saúde humana tem sido alvo de estudos recentes (KOSKINENA et al., 2017; MOISSL-EICHINGER et al., 2018).

1.1.1- *Halobacterium salinarum* NRC-1

H. salinarum NRC-1 é um microrganismo halófilo obrigatório, integrante do domínio Archaea. Foi isolado pela primeira vez em 1922 em peixes de água salgada (HARRISON; KENNEDY, 1922). Pode ser encontrado em salinas, lagos de água salgada ou até mesmo, alojado em cristais de sal (Figura 2)(DASSARMA, 2015). Apresenta crescimento ótimo em meio de cultura com 4,3 M de NaCl (concentração cerca de 10 vezes maior que a da água do mar) e temperatura de 37°C (KAUR et al., 2006) com tempo de geração aproximado entre 8-12 horas (DYALL-SMITH, 2008). Comparado a outros extremófilos do domínio Archaea, é um microrganismo de fácil cultivo em laboratório, atributo que o tornou elegível como organismo modelo desse domínio. Tem como características morfológicas a forma de bastonete e flagelos polares (DYALL-SMITH, 2008; OREN, AHARON, 2012). Possui características celulares distintivas como: vesículas de gás que permitem flutuação celular em uma coluna de água (PFEIFER et al., 1997) e a proteína de membrana bacteriorodopsina

ligada a um grupamento retinal que permite o uso da luz para a produção de ATP, sendo também um cromóforo que lhe confere a coloração púrpura (OESTERHELT; STOECKENIUS, 1971). Apresenta grande versatilidade metabólica sendo capaz de oxidar aminoácidos (SCHÄFER; PURSCHKE; SCHMIDT, 1996); respirar anaeróticamente usando aceptores de elétrons como: dimetil sulfoxido, nitrato e fumarato (OREN, AHARON; TRÜPER, 1990); fermentar arginina (RUEPP; SOPPA, 1996). Apresenta grande resistência a dessecação e a radiação UV (OREN, A., 2014).

Figura 2- *H. salinarum*. A) Microscopia eletrônica (aumento: 13.500x) evidenciando a forma de bastonetes e a presença de flagelos. (Fonte: https://www.biochem.mpg.de/522218/Org_Hasal). B) Cristais de sal colonizados por *H. salinarum*. Coloração rosa devido ao cromóforo retinal presente na membrana. (Fonte: (DASSARMA, 2015)). C) Microscopia eletrônica de *H. salinarum* (barra – 325nm), corte transversal. Na imagem é possível visualizar as vesículas de gás ocupando grande parte do citoplasma. (Fonte: DASSARMA et al., 2013).

A proteína bacteriorodopsina foi descoberta em *H. salinarum* em 1971 (OESTERHELT; STOECKENIUS, 1971), tal descoberta contribuiu para o estabelecimento desse organismo como modelo para estudos bioquímicos. Essa proteína de membrana de aproximadamente 27 Kda possui sete segmentos altamente hidrofóbicos que cruzam a bicamada lipídica (Figura 3). Possui um grupo retinal (lipídeo derivado da vitamina A) ligado por uma base de Schiff a um resíduo de lisina. Essa proteína funciona como uma bomba de prótons dependente de luz. Quando em contato com a luz, o grupo retinal, inicialmente em conformação *trans*, absorve um fóton e sofre uma fotoisomerização para a conformação *cis*, forçando uma mudança de conformação da proteína. Essa mudança de conformação leva ao distanciamento e variações no pKa de grupamentos químicos internos, causando a perda de

prótons de alguns resíduos de aminoácidos. Os prótons perdidos são lançados para a porção externa da membrana, de modo que, para recuperar a conformação inicial e mais estável da proteína, prótons são sequestrados da porção interna, criando um potencial elétrico de membrana que permite a volta dos prótons para a porção interna de forma energeticamente favorável através de um complexo ATP-sintase, servindo como via de produção de ATP na ausência de oxigênio (Figura 4). Devido à facilidade de sua imobilização, a bacteriorodopsina de *H. salinarum* foi usada para várias aplicações em estudos de modelos fotoquímicos (OREN, A, 2002). Além disso, a elucidação do mecanismo de produção de ATP em que está envolvida conferiu uma importante evidência que suporta o modelo de acoplamento quimiosmótico de Mitchell (RACKER; STOECKENIUS, 1974).

Figura 3- Bacteriorodopsina. (PDB ID: 2AT9). A cadeia polipeptídica única se dobra em sete alfa-hélices hidrofóbicas transversais à bicamada lipídica. O pigmento retinal se encontra em meio as hélices no interior da membrana. (Fonte: Lehninger Princípios de bioquímica, Sexta Edição. 2013).

Figura 4- Bomba de prótons. A interação com a luz leva à modificações estruturais no pigmento retinal, provocando alterações químicas em resíduos de aminoácidos internos da bacteriorodopsina que culmina no lançamento de prótons H^+ para fora da célula. Os prótons voltam para a célula a favor de um gradiente eletroquímico por uma ATPase, levando a produção de ATP. (Adaptado de: DASSARMA et al., 2013).

Outra característica marcante e de natureza proteica nesse organismo são as vesículas de gás (Figura 2C). Essas partículas proteicas cheias de gás são utilizadas para flutuação na coluna de água, permitindo às células localizar condições mais propícias de oxigenação e iluminação (PFEIFER et al., 1997). Quando presentes em grande quantidade, as vesículas de gás reduzem o volume do citoplasma, aumentando a proporção superfície/volume, o que pode ajudar as células a sobreviverem em condições de estresse (PFEIFER, 2012). A dispersão da luz pelas vesículas de gás influencia a coloração das colônias de *H. salinarum*. Em meio sólido, as células contendo vesículas formam colônias rosas e opacas (chamadas de Vac^+), enquanto as células sem vesículas formam colônias translúcidas de cor avermelhada (Vac^-) (Figura 5). A taxa de surgimento de mutantes espontâneos para vesículas de gás (Vac^-) é alta, aproximadamente 10^{-2} (LARSEN; OMANG; STEENSLAND, 1967). O surgimento de colônias Vac^- está muitas vezes relacionado a ação de elementos de transposição (insertion sequences IS) que interrompem genes ou regiões promotoras relacionados a produção das

vesículas (PFEIFER; GHARAMAN, 1993). O fenótipo facilmente detectado visualmente, somado as avaliações dos *loci* em que se encontravam IS, fornecem resultados que somado a anteriores, evidenciam que o genoma de *H. salinarum* apresenta considerável grau de instabilidade por conta das IS, principalmente em seus plasmídeos (LEIGH et al., 2011; SAPIENZA; ROSE; DOOLITTLE, 1982; SIMSEK et al., 1982). A formação de vesículas de gás em *H. salinarum* envolve 14 genes (gas vesicle protein - *gvp* genes), esses genes ficam agrupados em regiões genômicas chamadas de “*vac regions*” e são nomeados de *gvp* procedido pela letra correspondente a seu *locus* (e.g. *gvpA* e *gvpM*), a nomenclatura dos produtos proteicos desses genes segue a mesma lógica porém com letras maiúsculas (e.g. GVPA e GVPM)(PFEIFER et al., 1997). No genoma de *H. salinarum*, são encontradas duas *vac regions* (Figura 6). A regulação dos genes *gvp* é complexa, varia de acordo com situações de estresse e diferentes estágios durante o crescimento (PFEIFER, 2012). Um possível mecanismo de regulação é por meio de asRNAs codificados em *cis* ao gene *gvpD* (KRUGER; PFEIFER, 1996). Dois asRNAs complementares respectivamente as porções 5’ e 3’ desse gene foram detectados por *Notthern Blot* em momentos do crescimento em que os níveis da proteína codificada por esse gene estão baixos. Os autores do estudo especularam que os asRNAs podem estar atuando inibindo a tradução do mRNA ou criando um substrato dupla fita para a ação de uma RNase que degrada dsRNAs.

Por se tratarem de proteínas facilmente isoláveis, estáveis e que não causam resposta inflamatória em murinos, as proteínas GVPs de *H. salinarum* apresentam potencial uso terapêutico. Uma construção gênica em plasmídeo de superexpressão transformado em *H. salinarum* foi realizada para unir a instável proteína anti-inflamatória BPI (*Bactericidal/permeability-increasing protein*) às proteínas GVPs. A proteína quimérica resultante foi purificada e apresentou atividade anti-inflamatória e bactericida em ratos (BALAKRISHNAN et al., 2016).

Figura 5- Mutantes de vesículas de gás. Colônias de *H. salinarum* crescida em meio sólido. Ao centro, uma colônia avermelhada e translúcida, característica dos mutantes (Vac^-). Demais colônias rosas e opacas são linhagens sem mutações em genes de vesículas de gás (Vac^+). Fonte: (PFEIFER, 2015).

Figura 6- *vac* regions. Cada letra representa um *locus* gênico. Variações nos tamanhos dos *loci* podem ser observadas entre as *vac* regions. As setas indicam a direção em que os genes estão codificados. As diferentes *vac* regions são denominadas pelas letras *gvp1* e *gvp2*. O plasmídeo pNRC200 possui as duas regiões e o plasmídeo pNRC100 possui uma cópia da região *gvp1*. Adaptado de: (PFEIFER, 2012)

O genoma de *H. salinarum* NRC-1 foi sequenciado no ano de 2000, possui tamanho total de ~2,57 Mbp, dividido em três replicons circulares: um cromossomo (~2,01Mbp) e dois plasmídeos, pNRC100 (~191Kbp) e pNRC200 (~365Kbp) (NG et al., 2000). Na anotação mais recente do NCBI constam 2717 genes anotados, cerca de 25% desses genes estão anotados como codificantes de proteínas hipotéticas. O genoma apresenta alto conteúdo GC, aproximadamente 68% no cromossomo, característica provavelmente selecionada como

adaptação aos ambientes com alta incidência de luz UV em que esse organismo é encontrado naturalmente. Outra adaptação molecular evidenciada com o sequenciamento do genoma é o valor médio do pI (ponto isoelétrico) das proteínas, que é 5,1, indicativo de um proteoma altamente ácido, característica relacionada a adaptação às elevadas concentrações de sal em que sobrevivem. No genoma, foram anotadas 91 IS, 69 dessas distribuídas nos dois plasmídeos, uma evidência de que a instabilidade dos plasmídeos reportada anos antes do sequenciamento do genoma pode estar relacionada a IS. Além da linhagem NRC-1, a linhagem R1 também tem seu genoma sequenciado (PFEIFFER et al., 2008). Os dois genomas são altamente similares e as duas linhagens provavelmente se diferenciaram recentemente a partir da mesma espécie. Estudos já publicados do proteoma da linhagem R1 (PFEIFFER; OESTERHELT, 2015) fornecem valiosos dados que permitem grande aumento na precisão da anotação do genoma de ambas linhagens.

O primeiro método de produção de linhagens *knockout* no domínio Archaea foi desenvolvido com *H. salinarum* NRC-1 (PECK; DASSARMA; KREBS, 2000). Mais recentemente, nosso grupo de pesquisa também contribuiu nesse aspecto com a criação de vetores modulares para aplicações na engenharia genética desse organismo (SILVA-ROCHA et al., 2015).

Com o genoma sequenciado e a disponibilidade de ferramentas moleculares para manipulação genética, uma grande quantidade de estudos do transcrito em diversas condições ambientais e modificações genéticas foi possível, estabelecendo *H. salinarum* NRC-1 como organismo modelo para estudos de regulação gênica no domínio Archaea. O transcrito desse organismo já foi avaliado com variação de condições como: salinidade, temperatura, oxigenação, iluminação, radiação UV, concentração de ferro, cobre, zinco, manganês e níquel (BEER et al., 2014; BONNEAU et al., 2007; COKER et al., 2007; KAUR et al., 2006; SCHMID et al., 2007; WHITEHEAD et al., 2006). Além desses, há dados de transcrito de linhagens *knockout* de genes que codificam fatores de transcrição (TURKARSLAN et al., 2011) e também dados advindos de experimentos de imunoprecipitação em larga escala, permitindo a identificação de sítios de ligação de vários fatores de transcrição (FACCIOTTI et al., 2007). Dados do proteoma em várias condições experimentais também estão disponíveis para esse organismo (VAN et al., 2009).

A disponibilidade de uma grande quantidade e diversidade de dados em larga escala disponíveis faz de *H. salinarum* NRC-1 um importante organismo para estudos no campo da biologia sistêmica, onde a integração de informações obtidas a partir de diferentes experimentos e condições visa compreender os organismos por uma ótica sistêmica, prevendo suas respostas biológicas mediante condições impostas (GE; WALHOUT; VIDAL, 2003). Integrando uma grande quantidade de dados de expressão gênica de *H. salinarum* NRC-1, Bonneau et al. (2007) elaboraram um modelo global de regulação gênica que foi capaz de prever os níveis de transcrição de cerca de 80% dos genes em novas condições testadas para este microrganismo. Os autores discutem que situações em que não foi possível prever a resposta transcricional podem estar relacionadas a influência regulatória de elementos pouco estudados nesse organismo, como por exemplo, os ncRNAs (BONNEAU et al., 2007).

1.2- RNAs não-codificantes

Durante anos, o que ficou conhecido por “Dogma central da biologia molecular”, foi o modelo do qual muitas hipóteses foram deduzidas e testadas, impulsionando avanços no conhecimento sobre genética, biologia molecular e bioquímica da vida nesse planeta. Previamente idealizado por James Watson em 1952 (EDDY, 2001), esse modelo estabelece que a informação estocada pelas moléculas de DNA serve de molde para síntese de proteínas na tradução, cabendo às moléculas de RNA um papel intermediário como mensageiras da informação genética.

O contínuo avanço dos estudos da natureza bioquímica da conversão da informação genética em proteínas levou à descoberta de novas classes de RNAs além dos mensageiros (mRNAs), aumentando a relevância dessas moléculas nos estudos ao longo do tempo. Sem dados experimentais, Francis Crick havia predito a existência de “RNAs adaptadores”, possíveis intermediários entre as trincas do código genético e o aminoácido codificado (CRICK, 1958). Crick argumentava que RNAs teriam sido preferencialmente selecionados em detrimento de proteínas para o papel de adaptadores ao longo do processo evolutivo, porque eram suficientemente pequenos e poderiam estabelecer pareamentos específicos para o reconhecimento de moléculas de RNA. Posteriormente, os “adaptadores de RNA” de Crick foram experimentalmente observados, são o que hoje conhecemos como RNAs transportadores (tRNAs) (HOAGLAND et al., 1958). Mais uma importante função foi

associada aos RNAs estudos mostraram que além de sua composição proteica, os ribossomos são compostos por RNAs estáveis, dando origem a classe dos RNAs ribossomais (rRNAs) (BRENNER; JACOB; MESELSON, 1961; GROS et al., 1961). A obtenção de evidências experimentais de que RNAs constituem o centro catalítico dos complexos ribossomais viria a deixar claro a essencialidade dessas moléculas no contexto do “dogma central” (NOLLER, H., 1993; NOLLER, H F; CHAIRES, 1972). Assim, a função dos RNAs apenas como intermediários foi expandida, levando ao reconhecimento dessas moléculas como participantes diretas da síntese de proteínas.

A descoberta de novas classes e funções de RNAs foi muito além do que tange diretamente os processos de transcrição e tradução. O conhecimento de novos transcritos não codificantes com funções estruturais, catalíticas e/ou regulatórias como os tRNAs e rRNAs expandiu a compreensão do papel dos RNAs em vários processos celulares, principalmente na regulação da expressão gênica (EDDY, 2001). A identificação de novos RNAs não sobrepostos à regiões codificadoras têm contribuído para a compreensão da função das vastas regiões de DNA não-codificante dos genomas de eucariotos (EDDY, 2001; ST.LAURENT; WAHLESTEDT; KAPRANOV, 2015). Além disso, a compreensão do papel funcional de ncRNAs nas células têm levado ao desenvolvimento de novas técnicas que proporcionaram avanços na biologia molecular, como no silenciamento de genes por RNAi (REYNOLDS et al., 2004) e edição gênica com CRISPR (DOENCH, 2017).

São classificados como RNAs não-codificantes (ncRNAs) todos transcritos que não são mRNAs, ou seja, que não codificam proteínas (OLIVAS; MUHLRAD; PARKER, 1997). Além dos rRNAs e tRNAs, outros ncRNAs evidenciam a essencialidade dessas moléculas na manutenção dos circuitos celulares devido a sua função catalítica, conservação evolutiva e ubiquidade nos organismos. Podem ser citados como exemplo: RNA componente da ribozima RNase P, responsável pela maturação de extremidades 5' de tRNAs (GUERRIER-TAKADA et al., 1983); 4.5S RNA, componente da partícula de reconhecimento de sinal (*signal recognition particle* SRP) uma ribonucleoproteína envolvida no tráfego interno e secreção de proteínas nas células (ROSENBLAD et al., 2009); tmRNA (*transfer-messenger* RNA), componente do complexo ribonucleoproteico envolvido na liberação de ribossomos presos à “mRNAs problemáticos” (HAFEZ et al., 2013); TER (*Telomerase RNA*), componente das

telomerasas atuando como um molde para a síntese dos telômeros (WEBB; ZAKIAN, 2016). Tal versatilidade funcional e estrutural dos RNAs somada ao fato de estarem no cerne da maquinaria de tradução e fazerem parte de outros processos essenciais para a manutenção e reprodução celular, são alguns dos atributos que levaram pesquisadores a propor a hipótese de que essas moléculas podem ter sido os primeiros replicadores que surgiram na Terra, dando origem as formas de vida como conhecemos hoje (GILBERT, 1986).

O emprego das tecnologias de sequenciamento em larga escala no estudo dos transcritomas (e.g. RNA-seq e tiling array) propiciou significativos avanços na descoberta de novos ncRNAs em procariotos e eucariotos. A classificação dos ncRNAs é baseada em critérios como: tamanho, estrutura, localização em relação a um gene e mecanismos de atuação. O tamanho estimado dos ncRNAs é o atributo comumente usado para classificá-los em dois grandes grupos: longos ncRNAs (lncRNAs) quando maiores que 200 nucleotídeos (nt), ou pequenos (sRNAs do inglês *short RNAs*) quando menores que 200 nt (KAPRANOV et al., 2007; ST.LAURENT; WAHLESTEDT; KAPRANOV, 2015). ncRNAs também podem ser classificados em *cis* quando são expressos a partir do mesmo *locus* que seu alvo, compartilhando grande ou completa complementariedade com esse, ou em *trans*, quando expressos a partir de um *locus* diferente de seu alvo, compartilhando apenas complementariedade parcial. A quantidade de classes baseadas em outros critérios é numerosa e aumenta na medida em que as técnicas de sequenciamento, preparo de bibliotecas de RNA e ferramentas de predição *in silico* de maior resolução são desenvolvidas.

Apesar de lncRNAs também serem reportados em bactérias (BERNSTEIN et al., 1993) e vírus (REEVES et al., 2007), a literatura referente a essa classe em eucariotos é mais vasta, com várias caracterizações funcionais já realizadas (ANGRAND et al., 2015; KUNG; COLOGNORI; LEE, 2013; ZUR BRUEGGE; EINSPANIER; SHARBATI, 2017). São frequentemente encontrados no núcleo, podendo atuar em *cis* ou em *trans* e estão envolvidos em processos celulares como: organização da cromatina, enovelamento de proteínas, regulação gênica pré e pós-transcricional e podem atuar como “esponjas” para pequenos RNAs reguladores (RINN, 2015).

Em eucariotos, miRNAs (*micro RNAs*) e siRNA (*small interfering RNAs*) são exemplos de sRNAs bem conhecidos quanto a sua biogênese, mecanismos de atuação e seu

papel na regulação gênica. Ambos tem como característica em comum tamanho aproximado entre 20-30nt e atuam por pareamento de bases antisense com seus RNAs alvo, podendo bloquear a tradução ou levar à degradação. Estima-se que mais da metade dos mRNAs de uma célula eucariótica seja alvo de regulação de algum desses dois ncRNAs (FRIEDMAN et al., 2009).

Os miRNAs são transcritos como RNAs precursores de maior extensão (pré-miRNA) que podem ser codificados em regiões intrônicas. São processados por RNases tipo III (Drosha e Dicer) e incorporados a um “complexo silenciador” (HA; KIM, 2014). São reguladores pós-transcricionais que agem reprimindo a tradução de mRNAs alvo por pareamentos inespecíficos, geralmente com a região 3'-UTR (FRIEDMAN et al., 2009). Em humanos, a disfunção ou deficiência de miRNAs tem sido associada a doenças como o câncer (VANACORE et al., 2017). Além disso, miRNAs podem estar envolvidos na evolução da complexidade corporal dos eucariotos superiores (DOLGIN, 2012).

Os siRNAs provavelmente evoluíram como constituintes de um sistema de defesa celular contra vírus (HAASNOOT; WESTERHOUT; BERKHOUT, 2007). São oriundos de RNAs dupla fita (dsRNAs) processados pela enzima Dicer e incorporados a proteínas Argonautas, podendo assim, interagir com seus RNAs alvo caracterizando o mecanismo conhecido como RNAi (*interference RNA*) (FILIPOWICZ et al., 2005). siRNAs podem ser gerados a partir de RNAs exógenos ou endógenos, advindos de dsRNA que naturalmente ocorrem nas células (WATANABE et al., 2008). O silenciamento de genes por RNAi pode ser reproduzido artificialmente em laboratório e se tornou uma importante ferramenta para o estudo de funções gênicas e inibição de infecções virais (REYNOLDS et al., 2004).

Os snoRNAs (*small nucleolar RNAs*), também constituem um grupo de sRNAs bem conhecidos em eucariotos. Esses RNAs desempenham sua função no nucléolo e guiam modificações sítio específica em rRNAs (DECATUR et al., 2007). Baseados em motivos de sequências, são divididos em dois grupos: C/D box, envolvidos na metilação (normalmente inserção do grupamento 2'-O-metil em um nucleotídeo) e H/ACA box, envolvidos na pseudouridilação (substituição da ligação N-glicosídica por uma ligação carbono-carbono em uma uracila). Esses RNAs se associam a proteínas e formam complexos conhecidos como *small nucleolar ribonucleoprotein* (snoRNP), nos quais atuam como guias para as modificações nos RNAs alvo (BACHELLERIE; CAVAILLÉ; HÜTTENHOFER, 2002).

A maioria dos ncRNAs conhecidos em procariotos são classificados como sRNAs e assim como em eucariotos, desempenha sua função por pareamento de bases antisense com seus alvos, estando predominantemente envolvidos na regulação gênica. Apesar dessas semelhanças, a biogênese, proteínas com que atuam em conjunto e regiões alvo nos mRNAs são diferentes nos sRNAs desses dois grandes grupos (WAGNER, E GERHART H; ROMBY, 2015). Diferentemente dos bem caracterizados siRNA e miRNA, sRNAs procarióticos são comumente transcritos como entidades únicas e não passam por processamento (GOTTESMAN; STORZ, 2011).

Em bactérias, sRNAs estão envolvidos na regulação de diversos processos celulares como: homeostase de íons de ferro (SMALDONE et al., 2012); remodelamento de membrana (CHEN et al., 2004), motilidade e formação de biofilme (MIKA; HENGGE, 2013), regulação de transportadores (URBANOWSKI; STAUFFER; STAUFFER, 2000), regulação de fatores de transcrição (MANDIN; GUILLIER, 2013), metabolismo de açúcar e virulência (PAPENFORT; VOGEL, 2014). Os crRNAs (*Clustered Regularly Interspaced Short Palindromic Repeats RNAs– CRISPR RNAs*) são sRNAs descobertos em procariotos que merecem destaque pela atuação no sistema imune adaptativo baseado em RNA que age contra a invasão de DNAs exógenos advindos de plasmídeos conjugativos ou vírus (WIEDENHEFT; STERNBERG; DOUDNA, 2012). As regiões genômicas com sequências palindrômicas curtas espaçadas por sequências correspondentes a DNAs invasores foram descobertas em *Escherichia coli* em 1987 (JANSEN et al., 2002). Essas regiões são transcritas, processadas e dão origem a um crRNA, que é composto por um elemento espaçador transcrito flanqueado por sequências de RNA repetitivo que em conjunto com a proteína Cas, servirá de guia para o reconhecimento e degradação de DNAs invasores (HOCHSTRASSER; DOUDNA, 2015).

Para várias espécies de bactérias, a proteína Hfq é necessária para que ocorra o pareamento entre sRNAs e seus mRNA alvos. Essa proteína é uma chaperona de RNAs de estrutura homohexamérica pertencente a família Sm (compartilham o domínio Sm envolvido na ligação a RNAs) que pode catalisar o pareamento entre duas moléculas de RNA simplesmente por aproximá-los fisicamente, aumentando a velocidade de formação do produto, que nesse caso é um dsRNA. A formação de dsRNAs mediada pela proteína Hfq pode fazer parte de mecanismos de regulação pós-transcricional que envolvem modificações

nos sítios de ligação da maquinaria ribossomal ao mRNA, degradação ou aumento da estabilidade de RNAs (Figura 7) (VOGEL, JÖRG; LUISI, 2011).

Comparado a grande quantidade de estudos de ncRNAs em bactérias e eucariotos, esse campo de estudos ainda se encontra em estágio inicial no domínio *Archaea*, contendo poucas caracterizações funcionais para essas moléculas (GELSINGER; DIRUGGIERO, 2018a). O primeiro sRNA identificado no domínio *Archaea* foi um snoRNA C/D box. Mesmo sem a presença de um nucléolo nos organismos pertencentes ao terceiro domínio da vida, o nome desse sRNA foi mantido por conta da homologia de sequência em relação aos snoRNAs previamente identificados em eucariotos (DENNIS; OMER, 2005). Tal descoberta evidenciou que além de outras características moleculares unicamente compartilhadas por arqueias e eucariotos (EME et al., 2017), a conservação de ncRNAs também reflete a proximidade evolutiva desses dois grupos.

Figura 7- Possíveis modos de ação de Hfq. A) Hfq pode facilitar o pareamento de bases entre um sRNA e seu RNA alvo levando à oclusão do sítio de ligação da maquinaria ribossomal. B) A interação entre um sRNA e mRNA mediada por Hfq pode levar ao aumento da tradução por propiciar modificações estruturais no mRNA que facilitam o acesso da maquinaria ribossomal. C) A interação com Hfq pode proteger o sRNA de degradação por RNases. D) Indução da degradação por RNases. E) Hfq pode induzir o aumento da poliadenilação do mRNA por poli-A polimerases (PAP). Tal modificação, serve como sinal para degradação por exoribonucleases. Adaptado de (VOGEL; LUISI, 2011).

Após a descoberta dos snoRNAs, vários *cis* e *trans* sRNAs como RNAs derivados de fragmentos de tRNAs (trRNAs) e crRNAs foram identificados nesse domínio (BABSKI et al., 2014). O primeiro sRNA com seus alvos caracterizados experimentalmente em arqueias, o sRNA₁₆₂, foi identificado na arqueia metanogênica *Methanosarcina mazei* Gö1 combinando predições *in silico* com abordagens experimentais (PERNITZSCH et al., 2012; PRASSE; EHLERS; SCHMITZ, 2013). Esse sRNA atua simultaneamente em *cis* e em *trans*, interagindo com a região 5' de dois mRNAs, sobrepondo os sítios de interação com ribossomos (RBS) na região 5'UTR desses mRNAs. A função regulatória desse sRNA e das proteínas que seus mRNAs alvos codificam está relacionada ao intercâmbio entre o metabolismo de metanol e metilamina nesse organismo. Outro importante sRNA identificado como sRNA₁₅₄, também encontrado na arqueia metanogênica *M. mazei* Gö1, possui papel essencial no processo na fixação de nitrogênio, sendo caracterizado como um regulador chave no metabolismo desse organismo (EHLERS; JÄGER; SCHMITZ, 2011). Recentemente, um ncRNA que forma um dsRNA com um transcrito de sua própria fita antisense foi identificado em *Sulfolobus acidocaldarius*. Esse ncRNA está envolvido na regulação da formação de biofilme nesse organismo (ORELL et al., 2018).

Um destaque no que diz respeito a estudos de ncRNAs no terceiro domínio da vida é o halófilo *Haloferax volcanii*. Muitas abordagens *in silico* e *in vitro* tem sido utilizadas para a identificação dessas moléculas nesse organismo (BABSKI et al., 2014; GELSINGER; DIRUGGIERO, 2018a; KLIEMT; SOPPA, 2017; SOPPA et al., 2009). Há evidências de que mais de 40 sRNAs estão envolvidos na adaptação a variações de salinidade do meio em *H. volcanii*, indicando que essas moléculas podem ter um papel fundamental na regulação osmótica em halófilos (HEYER et al., 2012).

Nosso grupo de pesquisa também tem contribuído para a descoberta de novos ncRNAs em arqueias. Zaramela et al. (2014) mostraram que RNAs associados ao início de transcrição

(*Transcription start site associated RNAs* – TSSaRNAs) gerados provavelmente por pausas da RNA polimerase em *H. salinarum* NRC-1, são ubíquos nos três domínios da vida. No mesmo organismo, Gomes-Filho et al. (2015) encontraram RNAs sobrepostos (*sense overlapping transcripts* – sotRNAs) a região 3' de genes codificadores de transposases. A superexpressão de um desses sotRNAs gerou alterações fenotípicas evidenciando um possível papel funcional. Também para esse microrganismo, utilizando dados de tilling array, foram identificados 61 ncRNAs intergênicos diferencialmente expressos ao longo de uma curva de crescimento (KOIDE et al., 2009).

A interação de sRNAs com proteínas para exercer funções regulatórias como nos complexos de silenciamento em eucariotos e com a proteína Hfq em bactérias ainda é um fenômeno pouco conhecido em arqueias. A Proteína LSm, que assim como a Hfq é uma chaperona de RNAs pertencente a família Sm, é uma possível mediadora da interação entre sRNAs e seus RNAs alvo no terceiro domínio da vida (FISCHER et al., 2010). Essas duas proteínas apresentam uma alta conservação de estrutura terciária e quaternária (VOGEL, JÖRG; LUISI, 2011). Até o momento há dados de interação LSm-sRNAs publicados apenas para o organismo *H. volcanii*. Fischer et al. (2010), utilizando um ensaio de RIP-chip, mostraram que LSm interage com vários sRNAs nesse organismo. Além disso, nesse estudo os autores mostraram também que a deleção do gene que codifica para LSm leva a uma deficiência de crescimento em *H. volcanii*.

Figura 8- Transcritos sobrepostos em procariotos. As setas indicam a direção em que os genes e os transcritos se encontram. Em azul-claro as porções 5' e 3'UTRs. Exemplos de ncRNAs sobrepostos a genes codificantes nessa figura: TSSaRNA, transcritos associados ao início de transcrição; intraRNA, RNAs intragênicos; asRNA, transcritos antisense; seRNA, transcritos sense; sRNA, pequeno RNA intergênico. Adaptado de CATEN, 2017.

Em procariotos, vários estudos têm mostrado uma grande diversidade de transcritos sobrepostos sense e antisense (Figura 8), muitos desses funcionais, evidenciando uma grande versatilidade de regulação gênica por sRNAs em genomas compactos (KROGER et al., 2012). Grandes avanços nos estudos de transcritos sobrepostos em procariotos foram feitos com o emprego da técnica dRNA-seq (*differential RNA-seq*) (Figura 9) (SHARMA; VOGEL, 2014). Essa técnica possibilita a identificação em larga escala de inícios de transcrição (*Transcript start sites* – TSS) ao longo do genoma de procariotos e de organelas eucarióticas com resolução de um único nucleotídeo. A técnica consiste na comparação de amostras de RNA tratadas com a enzima *5' Terminator Exonuclease* (TEX) com amostras que não receberam esse tratamento. Essa enzima degrada RNAs que apresentam apenas um grupo fosfato na extremidade 5', ou seja, transcritos que tenham passado por alguma etapa de processamento. Dessa forma, ocorre o enriquecimento da amostra tratada com transcritos primários. São preparadas bibliotecas para RNA-seq com as amostras tratadas e não tratadas enzimaticamente com TEX e após o sequenciamento, a comparação entre essas bibliotecas permite a identificação de TSS. O mapeamento de TSS dentro de regiões codificadoras ou em regiões intergênicas seguido de análises de sequências promotoras e dinâmica de expressão em diferentes condições fornece uma ampla perspectiva a respeito da localização genômica e funcionamentos dos ncRNAs em um organismo. Existem disponíveis dados de dRNA-seq para mais de 30 organismos, dentre esses, quatro arqueias: *H. volcanii* (BABSKI et al., 2016), *Methanobrevibacter smithii* (LI et al., 2015), *Thermococcus kodakarensis* (JAGER et al., 2014), *Thermococcus onnurineus* NA1 (CHO et al., 2017)) e *Methanocaldococcus jannaschii* DSM, constituindo um importante conjunto de dados para o avanço dos estudos sobre ncRNAs no terceiro domínio da vida.

O fato de que a maioria das bases que compõem os genomas de eucariotos e procariotos são transcritas em algum momento, fenômeno conhecido como “transcrição pervasiva” (WADE; GRAINGER, 2014), pode ser encarado por duas perspectivas: existe uma grande quantidade de ncRNAs funcionais a serem descobertos e esses compõem elementos que contribuirão para o aumento da nossa capacidade de predição em relação as complexas redes de regulação gênica (STEENSELS; VERSTREPEN, 2016); grande parte desses transcritos constituem “ruídos” relacionados a erros do processo de transcrição ou artefatos técnicos (LLORENS-RICO et al., 2016; VAN BAKEL et al., 2010). Uma outra proposta para explicar

esse fenômeno é que a transcrição pervasiva permite uma atividade de controle constante de dano no DNA. Por meio de uma atividade de reparo acoplada, a RNA polimerase recruta a maquinaria de reparo quando necessário. Dessa forma, a transcrição frequente das duas fitas pode garantir maior integridade do genoma (KAMARTHAPU; NUDLER, 2016). Se a maioria dos transcritos detectados com métodos de larga escala são funcionais ou não, é um tema que ainda causa muitas discussões, com evidências apresentadas pelos dois lados. Porém, não resta dúvidas sobre a importância dos ncRNAs nos processos celulares e de que a identificação dessas moléculas e de suas funções são passos na direção de uma melhor compreensão dos mecanismos de regulação gênica. A integração de dados em larga escala de alta resolução, como o dRNA-seq, com outros dados, permite uma visão ampla do papel dos sRNAs nas células. O mapeamento de ncRNAs é um primeiro passo essencial na descoberta de novos RNAs reguladores.

Figura 9- dRNA-seq. A) Previamente ao sequenciamento (RNA-seq), uma das amostras é tratada com a enzima TEX, causando o enriquecimento de transcritos primários na amostra. B) As bibliotecas sequenciadas são alinhadas e o acúmulo estatisticamente significativo de *reads* em posições específicas da biblioteca tratada em relação a não tratada permite a identificação de TSS. C) Os TSS são comumente classificados em: primários (pTSS), mapeados no início de gene; internos (iTSS), mapeados na mesma fita e dentro de regiões com genes anotados; órfãos (oTSS), mapeados em regiões intergênicas; antisense (aTSS), mapeados dentro de genes anotados, porém na fita oposta. Adaptado de (BISCHLER et al., 2015)

1.3- RNAs antisense

RNAs antisense (asRNAs) são ncRNAs transcritos em *cis* na fita oposta a um gene, podendo o último ser codificador de proteína ou não (PELECHANO; STEINMETZ, 2013). A forma como esses ncRNAs estão codificados no genoma sugere seu modo de ação em *cis* por conta do favorecimento termodinâmico natural devido ao pareamento de bases completo com o mRNA de seu gene cognato e proximidade física. O pareamento de bases antisense de um ncRNA com seus RNAs alvo é um fenômeno predominante na ação dessas moléculas, porém o que agrupa os asRNAs como uma classe é serem codificados em *cis* em relação a um gene, o que não restringe sua ação reguladora ao gene cognato, podendo também atuar em *trans* (SAYED; JOUSSELIN; FELDEN, 2012).

A transcrição simultânea de duas fitas opostas de DNA da mesma região foi reportada pela primeira vez em 1969, em um estudo com fago lambda (BOVRE; SZYBALSKI, 1969). Inouye (1988) listou em sua revisão o que chamou de “todos os asRNAs de ocorrência natural conhecidos”. Essa lista é composta de 11 asRNAs, nove envolvidos na transcrição (reguladores negativos da expressão gênica de seus alvos) e dois na replicação de DNA. Todos RNAs listados eram de bactérias, e com apenas uma exceção, todos associados a funções não cromossomais como: controle de réplica de plasmídeos, repressão de fagos e transposases (INOUE, 1988). Não tardou para que asRNAs fossem encontrados em eucariotos, evidenciando a ubiquidade dessa classe de ncRNAs nos organismos vivos (VANHÉE-BROSSOLLET; VAQUERO, 1998).

O emprego das técnicas microarray e RNA-seq na análise de transcritomas em bactérias, arqueias e eucariotos mostrou que a transcrição de regiões sobrepostas é um fenômeno comum que gera uma grande quantidade de transcritos antisense a genes codificadores de proteínas (BEITER et al., 2009; GEORG; HESS, 2011; WURTZEL et al., 2010a). Cerca de 30% de todos genes anotados no genoma humano possuem pelo menos um asRNA associado (OZSOLAK et al., 2010). Em bactérias, a porcentagem de genes

codificadores de proteínas com transcritos antisense reportados varia de cerca de 2% em *Salmonella*, até cerca de 75% em linhagens de *Prochlorococcus* (WAGNER, E GERHART H; ROMBY, 2015), e mais recente, 80% para *Pseudomonas aeruginosa* (ECKWEILER; HÄUSSLER, 2018).

Inicialmente, a hipótese de que a transcrição antisense é pervasiva, caracterizando se como regra e não exceção foi encarada com ceticismo. Argumentos contra essa proposta eram direcionados às dificuldades técnicas de manter a informação de especificidade da fita dos transcritos em abordagens de larga escala, o que levava a uma grande possibilidade de ocorrência de artefatos técnicos (JOHNSON et al., 2005). Grandes avanços ocorreram nesse quesito, os protocolos para manutenção da informação de fita de origem dos transcritos evoluíram, aumentando significativamente a resolução dos métodos de análise de transcritomas, proporcionando maior confiabilidade na inferência de fenômenos biológicos e descobertas de ncRNAs relacionados a transcrição antisense (LASA, INIGO; TOLEDO-ARANA; GINGERAS, 2012; LEVIN et al., 2010; LYBECKER et al., 2014; PELECHANO; STEINMETZ, 2013; SHARMA; VOGEL, 2014; SUN et al., 2017; WADE; GRAINGER, 2014).

Transcritos antisense podem exercer sua função de reguladores em diferentes estágios da expressão gênica, sendo que, apenas a transcrição de uma região antisense e não necessariamente a molécula gerada pode promover a regulação. A interferência transcricional pode ocorrer previamente ao início da transcrição do gene regulado, de modo que, o processo de transcrição de uma das entidades do par sense/antisense impede que seu cognato seja transcrito, ou seja, um processo de transcrição inibe outro (SHEARWIN; CALLEN; EGAN, 2010). Tal fenômeno pode ocorrer por competição de regiões promotoras quando a montagem da maquinaria de transcrição em uma região do cromossomo impede fisicamente a montagem da outra necessária para a transcrição do gene cognato. Outra forma de interferência transcricional do gene cognato é a oclusão de regiões promotoras pela passagem constante da RNA polimerase pela região devido a expressão do *locus* antisense. Em eucariotos, são propostos mecanismos de inibição de transcrição por asRNAs envolvidos na metilação de ilhas CpG (LISTER et al., 2009) e remodelamento de cromatina (YU et al., 2008), ambos relacionados a eventos epigenéticos.

Figura 10- Efeitos co-transcricionais da transcrição antisense. A) A transcrição convergente de dois transcritos em fitas opostas pode levar ao choque de RNAs polimerases e término prematuro dos transcritos. B) A transcrição de um *locus* antisense pode bloquear sítios de *splicing* determinando a isoforma do mRNA. C) O tamanho de um transcrito pode ser determinado por interferência transcricional promovida por transcrição antisense. Adaptado de: PELECHANO; STEINMETZ, 2013.

Posteriormente ao início do processo de transcrição, a regulação por transcritos antisense ou interferência transcricional antisense podem ocorrer co-transcricionalmente. A colisão de RNAs polimerases pode ocorrer quando duas fitas de uma mesma região estão sendo transcritas simultaneamente, levando ao não alongamento completo de um ou ambos transcritos (Figura 10A) (HOBSON et al., 2012). Ainda no contexto co-transcricional, a transcrição antisense pode determinar qual isoforma de transcrito será produzida pelo gene cognato. Isso pode ocorrer quando o asRNA bloqueia o sítio de *splicing* do mRNA evitando seu processamento (Figura 10B). O controle de isoformas mediado por asRNAs também acontece em procariotos. O fenômeno conhecido como “atenuação da transcrição” (YANOFSKY, 1988), pode ser mediado por asRNAs afetando o tamanho do mRNA (Figura 10C) (PELECHANO; STEINMETZ, 2013). Na bactéria *Vibrio anguillarum*, o asRNA de 427 nt chamado de *RNAB* codificado em *cis* a um *operon* de genes que codificam proteínas

transportadoras de ferro, atua como terminador prematuro da transcrição desse mesmo *operon*, determinando o tamanho do mRNA policistrônico (STORK et al., 2007).

Figura 11- Regulação pós-transcricional por asRNAs. A) asRNA oclui o sítio de ligação de um miRNA evitando a degradação do mRNA. B) TAs do tipo I em *E. coli*. A ligação do asRNA *SymR* ao mRNA do gene *SymE* impede a ligação do complexo ribossomal ao mRNA. C) Em bactérias, dsRNAs podem ser degradados pela RNase III. D) Transcritos antisense (vermelho) induzem um limiar para a expressão de seu gene cognato (azul). Genes sem transcritos antisense associados seguiriam uma cinética de expressão diferente (cinza). Adaptado de: PELECHANO; STEINMETZ, 2013.

Na regulação pós-transcricional é necessário que ambos transcritos, sense e antisense, estejam presentes. A transcrição antisense por si só não caracteriza mais um possível fenômeno regulador nesse estágio, logo, é necessário um asRNA pronto para atuar. Como característica inerente a todos processos de regulação pós-transcricional, a regulação por asRNAs nesse estágio permite respostas celulares mais rápidas do que as regulações nos estágios citados anteriormente. Tem como alvo os mRNAs já presentes no citoplasma e salta a fase *lag* entre a mudança nos níveis de transcrição e o estabelecimento de novas concentrações de mRNAs de proteínas reguladoras que caracterizarão uma resposta. asRNAs

podem atuar diminuindo ou aumentando a estabilidade e a frequência de tradução de um mRNA (PELECHANO; STEINMETZ, 2013).

Um asRNA pode aumentar a estabilidade de um transcrito ocluindo sítios que seriam alvos para processos de degradação. Um exemplo ocorre em humanos, onde um asRNA forma um dsRNA com o mRNA do gene que codifica uma secretase com papel-chave na progressão da doença de Alzheimer (FAGHIHI et al., 2010). A formação desse dsRNA bloqueia sítios de ligação para um miRNA (Figura 11A). A formação de um dsRNA também pode bloquear sítios de reconhecimento para a RNase E, que cliva RNAs fita simples, diminuindo a taxa de degradação do transcrito (STAZIC; LINDELL; STEGLICH, 2011).

Em bactérias, dsRNAs são degradados por RNases do tipo III, sugerindo a atuação de asRNAs como reguladores negativos, marcando seus RNAs alvo para degradação (Figura 11C). Lasa et al. (2011) evidenciaram a ação da RNase III sequenciando bibliotecas de RNAs de diferentes tamanhos e utilizando linhagens *knockout* do gene que codifica RNase III. O processamento de transcritos sobrepostos resultando em pequenos RNAs de cerca de 22 nt é dependente da RNase III e cerca de três quartos dos RNAs pertencentes a genes anotados passam por processamento mediado por asRNAs e RNase III em *Staphylococcus aureus* (LASA et al., 2011). Lybecker et al. (2014) utilizaram anticorpos monoclonais específicos para dsRNAs em linhagens de *E. coli knockout* do gene da RNase III para estudar esse fenômeno. Os autores realizaram imunoprecipitações procedidas de RNA-seq e compararam os resultados entre linhagens com e sem o gene da RNase III. Foram encontrados 316 dsRNAs contra 31 na linhagem selvagem, fornecendo mais uma evidência desse mecanismo em bactérias (LYBECKER et al., 2014).

No contexto de asRNAs servindo como sinalizadores para degradação após formar um dsRNA, é postulado que essas moléculas impõem um limiar no qual apenas os mRNAs que superam a concentração de seus pares antisense escaparão da degradação (Figura 11D). Transcritos espúrios seriam neutralizados por asRNAs, evitando qualquer tipo de descontrole que poderiam causar na célula (LASA, et al., 2011; PELECHANO; STEINMETZ, 2013).

Alterações conformacionais promovidas pela ligação de um asRNA à extremidade 5' de um transcrito podem aumentar a eficiência de ligação e/ou reconhecimento dos ribossomos, aumentando a tradução do mRNA (CARRIERI et al., 2012). Inversamente, o pareamento de um asRNA na região da extremidade 5' do mRNA pode inibir a ligação dos

ribossomos ao transcrito, inibindo sua tradução. Em bactérias, a sequência conhecida como Shine-Dalgarno (SD), localizada *upstream* ao códon de início de tradução (SHINE; DALGARNO, 1974), possui papel fundamental no reconhecimento do mRNA pela subunidade ribossomal 30S; essa sequência rica em purinas é complementar a sequência anti-SD localizada na extremidade 3' do rRNA 16S (RINGQUIST et al., 1993). A maioria dos sRNAs bacterianos tem como alvo essa região, competindo de forma eficiente contra a ligação da subunidade 30S ao mRNA (WAGNER; GERHART; ROMBY, 2015). Um exemplo dessa regulação negativa é a mediada pelo asRNA *symR*, de 77 nt, que regula seu gene cognato *symE*, constituindo um sistema toxina-antitoxina (TAs) do Tipo I (Figura 11B) (KAWANO; ARAVIND; STORZ, 2007).

TAs são encontrados em procariotos, são compostos por pares de genes codificados em *operons* e regulados conjuntamente. Estão relacionados à persistência e dormência celular em bactérias (KEREN et al., 2004). Nesses sistemas, a toxina é caracterizada por proteínas tóxicas para as células que quando presentes podem interferir em processos essenciais como tradução e replicação ou na estabilidade da membrana celular (HARMS et al., 2018). A antitoxina codifica um produto instável (ncRNA ou proteína) que atua neutralizando a toxina. Em condições de estresse, as antitoxinas podem ser seletivamente inibidas, permitindo a ação das toxinas (HAYES; VAN MELDEREN, 2011). Três principais tipos de TAs são conhecidos: Tipo I (T: proteína A: asRNA), Tipo II (T: proteína A: proteína) e Tipo III (T: proteína A: ncRNA) (Figura 12) (CORAY et al., 2017). Os TAs do tipo I representam um exemplo de regulação pós-transcricional por asRNA bem compreendido funcionalmente e conservado em bactérias (HAYES; VAN MELDEREN, 2011).

A regulação gênica por asRNAs pode surgir pela forma como os genes estão organizados no genoma (SESTO et al., 2012). Mecanismos regulatórios da expressão gênica que ocorrem devido à proximidade com que se encontram os genes é uma característica de procariotos (e.g. *operons*). Em bactérias, há casos de asRNAs resultantes de 5' e 3' UTR longos, fazendo com que uma porção do mRNA seja um possível regulador antisense (Figura 13A) (HERNÁNDEZ et al., 2006; RASMUSSEN; NIELSEN; JARMER, 2009). Também há casos de lasRNAs (*long antisense RNAs*) que se estendem por *operons* inteiros e desempenham funções regulatórias (Figura 13B) (TOLEDO-ARANA et al., 2009). No contexto dos *operons*, um padrão recorrente de codificação de lasRNAs no qual a transcrição

do RNA começa antisense a um gene e continua incorporando outros genes na mesma fita é reportado em bactérias. Nesse padrão, o lasRNA possui uma região 5' que pode atuar como um regulador antisense e uma porção distal que atua como mRNA. Essa organização gênica recebeu o nome de “*excludon*” (Figura 13C) (SESTO et al., 2012).

Figura 12- Sistemas Toxina-antitoxina. Um asRNA atua como antitoxina e inibe a tradução do mRNA da toxina no Tipo I. No tipo II, toxina e antitoxina são traduzidas, a interação das duas proteínas neutraliza a toxina. Um ncRNA não sobreposto ao gene da toxina atua como antitoxina no Tipo III. A interação desse ncRNA com a proteína o efeito da toxina. Adaptado de (MRUK; KOBAYASHI, 2014).

A regulação gênica mediada pelo pareamento com um ncRNA regulador seguida de degradação conta com uma RNase capaz de reconhecer e clivar dsRNAs em eucariotos e bactérias (e.g. Dicer, Drosha e RNase III). Essas RNases possuem um motivo que permite a ligação a dsRNAs (dsRBM – *double strand RNA binding motif*) e são integrantes da família proteica “RNase III” (NICHOLSON, 2014). A ocorrência dessas proteínas é exceção no domínio *Archaea*, nos poucos organismos em que são encontradas, seus genes codificadores provavelmente foram obtidos por transferência horizontal de bactérias (NICHOLSON, 2014;

TIAN et al., 2004). Provavelmente, a degradação de dsRNAs nesses organismos ocorre por meio da ação cooperativa de várias enzimas com atividades catalíticas complementares ou por uma RNase alternativa, cuja função de degradação de dsRNAs ainda não foi elucidada (NICHOLSON, 2014).

Figura 13- lasRNAs. As caixas com a letra P se referem a regiões promotoras das Orfs (*open reading frames*) e dos asRNAs. A) Longos 5' e 3'UTR de mRNAs podem atuar como asRNAs de mRNAs de genes adjacentes. B) lasRNAs podem sobrepor operons inteiros. C) Excludon. Um lasRNA atuando também como um longo mRNA. Adaptado de (SESTO et al., 2012).

O primeiro relato de regulação da expressão gênica por pareamento antisense de um RNA em arqueias foi no ano de 1993, em *Halobacterium salinarum* (STOLT, PELTE; ZILLIG, 1993a; WAGNER, E G H; SIMONS, 1994). Um transcrito de 151 nt produzido por esse organismo é complementar a região 5' de um transcrito essencial para a replicação de um fago, reduzindo os níveis desse último. Os autores conseguiram detectar a formação do dsRNA e o fenômeno de degradação do mesmo, porém a enzima responsável pela degradação não foi identificada (STOLT, PELLE; ZILLIG, 1993b). Recentes estudos baseados em métodos de sequenciamento em larga escala, como o dRNA-seq, mostram que asRNAs provavelmente compreendem o maior grupo de ncRNAs nesse domínio, porém, poucas

caracterizações de mecanismos de regulação e seus alvos foram realizadas (GELSINGER; DIRUGGIERO, 2018a).

Figura 14- Correlação inversa dos níveis de transcritos sense e antisense em *H. volcanii*. Gráfico de dispersão da cobertura média por *reads* dos transcritos sense e antisense. Fonte: (BABSKI et al., 2016).

Estudos de asRNAs no halófilo *H. volcanii* constituem uma fonte de informações importante para a compreensão geral dessas moléculas no terceiro domínio da vida (GELSINGER; DIRUGGIERO, 2018a). Para esse organismo, asRNAs estão presentes em genes envolvidos em processos como: sinalização de quimiotaxia, regulação da transcrição e mobilidade de elementos de transposição (GELSINGER; DIRUGGIERO, 2018b) Utilizando a técnica dRNA-seq, foram encontrados 1244 aTSSs em *H. volcanii* (BABSKI et al., 2016). Os níveis de expressão dos asRNAs relacionados aos aTSSs encontrados em relação aos seus mRNAs cognatos apresentaram forte correlação negativa (Figura 14). Esse resultado pode ser uma indicação do papel de reguladores negativos dos asRNAs encontrados em *H. volcanii* por meio da formação de duplex asRNA/mRNA seguida de degradação, o que requer um mecanismo de degradação de dsRNA por RNases, algo ainda desconhecido *in vivo* nesse organismo (KLIEMT; SOPPA, 2017).

Há dois relatos de RNases oriundas de arqueias com atividade de degradação de dsRNAs *in vitro*. RNases H são conhecidas por sua atividade de degradação de moléculas de

RNA em híbridos RNA/DNA (TADOKORO; KANAYA, 2009). Uma RNase H do tipo I foi identificada e purificada na arqueia *Sulfolobus tokodaii* 7, essa enzima apresentou atividade de degradação de dsRNAs *in vitro* (OHTANI et al., 2004a). Porém, essa função não pôde ser generalizada para outras arqueias que possuem essa RNase codificada em seu genoma. Uma RNase homóloga a essa última também foi purificada de *H. salinarum* NRC-1, e nenhuma atividade de degradação de dsRNA foi detectada *in vitro* (OHTANI et al., 2004b).

As RNases R pertencem a família de proteínas *RNase II* da qual seus constituintes estão presentes nos três domínios da vida e desempenham papéis essenciais no metabolismo de RNA (ZUO; DEUTSCHER, 2001). RNases R possuem atividade exoribocatalítica, degradando RNAs fita simples na direção 3'-5' (VINCENT; DEUTSCHER, 2009). Além dessa atividade enzimática, também possuem papel central na degradação de RNAs estruturados e há relatos de atividade de degradação de dsRNAs (CHENG; DEUTSCHER, 2005; MATOS, RUTE GONÇALVES; BARBAS; ARRAIANO, 2009). Postula-se que sejam essenciais nas pouco compreendidas vias de degradação de RNAs em arqueias por ser a única exoribonuclase codificada no genoma de várias arqueias (PORTNOY; SCHUSTER, 2006). A única RNase R codificada no genoma do halófilo *H. volcanii* foi purificada e apresentou atividade de degradação de dsRNAs *in vitro* (MATOS, RUTE et al., 2012). O estudo do papel dessa RNase *in vivo* em *H. volcanii* é limitado pelo fato de não ser possível obter linhagens *knockout* viáveis do gene que codifica essa proteína (PORTNOY; SCHUSTER, 2006), o que leva a inferir que essa enzima é essencial para essa arqueia. Esses são os únicos estudo abordando o papel funcional de RNases R em arqueias.

De modo geral, a transcrição antisense é um fenômeno pervasivo e há claras evidências de que vários dos asRNAs oriundos desse processo são importantes moléculas reguladoras em eucariotos e bactérias. Há poucos estudos sobre o papel dos asRNAs e impacto das RNases na regulação gênica de arqueias, sendo que, os mecanismos de degradação de dsRNAs, importante forma de regulação que envolve asRNAs, permanecem desconhecidos no terceiro domínio da vida.

2-Objetivo Geral

Caracterizar o transcrito antisense primário de *H. salinarum* NRC-1 utilizando dados de dRNA-seq e dados de bibliotecas de RNA-seq *paired-end* para anotar asRNAs associados aos inícios de transcrição mapeados.

2.1- Objetivos específicos

- Identificar asRNAs possivelmente interagindo com a proteína LSm utilizando dados de RIP-seq;
- Verificar a presença de aTSSs em sistemas toxina-antitoxina do tipo II em arqueias utilizando dados de dRNA-seq.
- Avaliar os asRNAs diferencialmente expressos ao longo da curva de crescimento utilizando dados de RNA-seq.
- Identificar aTSSs possivelmente conservados entre *H. salinarum* NRC-1 e *H. volcanii*;
- Avaliar o efeito da deleção da RNase R nos asRNAs anotados utilizando RNA-seq de linhagem *knockout*.

3-Materiais e métodos

3.1-dRNA-seq

Todo o experimento de dRNA-seq do qual os resultados foram utilizados nesse trabalho foram realizados pelo Dr. Felipe ten Caten e serão resumidamente descritos abaixo para melhor compreensão dos dados que foram utilizados.

3.1.1- Linhagem e condição de cultivo

Pellets da linhagem selvagem *H. salinarum* NRC-1 em condição referência foram coletados como descrito na seção. Para a coleta de amostras ao longo da curva de crescimento, a linhagem NRC-1 foi cultivada em meio CM em volume de 200 ml à 37° C sob agitação (125 rpm) e luminosidade constantes. Os *pellets* foram coletados como previamente apresentado em três momentos distintos do crescimento, a 17 horas, 37 horas e 86 horas. Todos os experimentos foram realizados com réplica biológica.

3.1.2-Extração RNA dRNA-seq

A extração de RNA para o experimento de dRNA-seq foi feita utilizando o mesmo protocolo descrito na seção. Nesse experimento foi utilizado apenas a linhagem selvagem NRC-1 e as amostras foram extraídas a partir de *pellets* coletados em condição referência e em três pontos distintos de uma curva de crescimento como descrito na seção 3.17.

3.1.3- Depleção de RNAs processados

As amostras de RNA de cada réplica foram divididas em duas alíquotas. Uma delas foi tratada com 1U da enzima TEX (Epicentre TER51020) para cada 1µg de RNA (TEX+) enquanto a segunda alíquota foi incubada apenas na presença do tampão (TEX-). As reações foram incubadas a 30° C por 60 minutos e logo após foi realizada a purificação com o kit RNeasy MinElute Cleanup (Quiagen). Para verificar se houve degradação dos RNAs, processados (rRNAs) 1µL das amostras TEX+ e TEX-foi analisada em eletroforese em gel de agarose 1% com tampão de corrida TAE. Após a confirmação da depleção dos RNA processados as amostras foram tratadas com a enzima TAP (Tobacco Acid Pyrophosphatase) (Epicentre T19100) na concentração de 1U para cada 1µg de RNA a 37 °C por 30 minutos e purificadas novamente com o kit RNeasy MinElute Cleanup (Qiagen).

3.1.4- Preparação das bibliotecas de dRNA-seq

As bibliotecas para o experimento de dRNA-seq foram preparadas e sequenciadas seguindo o mesmo protocolo já descrito na seção anterior.

3.2-Processamento dos dados de RNA-seq

A qualidade dos dados brutos de RNA-seq foi inspecionada utilizando o programa FastQC v0.11.7 (ANDREWS, 2018). As bibliotecas *paired-end* foram processadas utilizando Trimmomatic v0.36 (BOLGER; LOHSE; USADEL, 2014) para trimagem dos adaptadores e sequências de baixa qualidade. Reads que passaram na trimagem como pares R1-R2 foram alinhados ao genoma de referência de *H. salinarum* NRC-1 utilizando o alinhador HISAT2 v2.1.0. (KIM, DAEHWAN; LANGMEAD; SALZBERG, 2015) no modo *paired-end*, permitindo uma distância limite de 1000nt entre os pares. *Reads* órfãos R1 e R2 que passaram pela trimagem pelo critério de qualidade, mas não como pares R1-R2, foram alinhados usando o modo *single-end*. Foi permitido o alinhamento de reads em mais de uma posição no genoma. Os arquivos SAM gerados no alinhamento foram convertidos para BAM utilizando SAMtools v1.3.1 (LI, HENG et al., 2009) e foram utilizados como entrada para o programa NMR para que a melhor posição de alinhamento dos reads mapeados em mais de uma posição fosse encontrada levando em conta a minimização da variância local da cobertura. A cobertura total do genoma foi computada utilizando deepTools v2.5.3 (RAMÍREZ et al., 2016) levando em consideração a extensão dos fragmentos inteiros para os alinhamentos *paired-end*. Para estabelecer o perfil de acúmulo de regiões 3' dos fragmentos sequenciados utilizando os reads R2 ao longo do genoma, foi utilizado bedtools v2.2.26 (QUINLAN, 2012). Os dados foram visualizados utilizando Gaggie Genome Browser (BARE et al., 2010).

3.3-Mapeamento de TSS em *H. salinarum* NRC-1

Para determinar de forma automatizada os inícios de transcrição de *H. salinarum* NRC-1, os dados de dRNA-seq foram analisados com a ferramenta TSSAR (AMMAN et al., 2014). Brevemente, a ferramenta analisa o enriquecimento de contagem de início de *reads* na biblioteca TEX+ em comparação a biblioteca TEX-, levando em consideração o nível de expressão da região na qual a posição analisada se localiza. Tomando o modelo estatístico adotado pelo programa, caso essa diferença seja maior do que o esperado a posição é determinada como um TSS (AMMAN et al., 2014). Para a determinação dos TSSs, foi

considerado um valor $p < 0,005$ como corte de significância estatística, exigido um mínimo de 4 *reads* na posição e uma distância de agrupamento de TSSs de pelo menos 5 nt.

3.4-Filtragem de aTSS em *H. salinarum* NRC-1

Para determinar os aTSSs, foi utilizada como referência a anotação do genoma de *H. salinarum* NRC-1 disponível no RefSeq atualizada em 2017 pelo *Prokaryotic Genome Annotation*. Adicionalmente, a anotação curada manualmente da linhagem *H. salinarum* R1 foi utilizada como complemento (PFEIFFER; OESTERHELT, 2015) considerando as CDS presentes em R1, mas ausentes na anotação de NRC-1. Utilizando *scripts* elaborados na linguagem R, todos TSSs localizados internos a regiões gênicas, porém localizados na fita oposta, foram classificados como aTSSs. Também foram filtrados todos TSSs localizados na fita oposta e com até 200 nt *downstream* a uma região gênica e não sobrepostos a um gene anotado. Quando um sinal de expressão gênica definido e sobreposto ao gene *upstream* estava associado a esses TSS, eles foram classificados como *downstream antisense* TSS (daTSS).

Potenciais estruturas secundárias relacionadas aos aTSSs foram filtradas após a avaliação do valor MFE (*minimum free energy*) de suas regiões *downstream*. Um script elaborado em nosso laboratório percorre o genoma extraindo subsequências de tamanho 25+1+25 com um deslocamento de 10nt levando a uma sobreposição de 41 nt entre as subsequências adjacentes. Todas subsequências extraídas foram submetidas ao programa RNAfold 2.0.5 (LORENZ et al., 2011) para predição de estruturas secundárias. Utilizando parâmetros em modo *default*. A distribuição dos valores de MFE obtidas para todas essas subsequências foi comparada com as sequências correspondentes as regiões imediatamente *downstream* aos aTSSs. O quantil 33% da distribuição de valores MFE de todo genoma foi escolhido como valor limiar ao qual aTSSs com regiões *downstream* de valor menor foram considerados estruturados e classificados como falsos positivos (Figura 15).

Figura 15- Distribuição de valores de MFE calculados para sequências de 51nt extraídas a cada 10nt (janela deslizando) ao longo do genoma de *H. salinarum* NRC-1. Em preto distribuição de valores para todo genoma. Em verde distribuição de valores para as sequências *downstream* aos aTSSs. Em vermelho delimitação do quantil de 33,3%. aTSSs com sequências *downstream* apresentando valores MFE abaixo desse quantil foram considerados possíveis falsos positivos.

3.5-Anotação dos *loci* de asRNAs

Para a anotação dos *loci* de asRNAs no genoma de *H. salinarum* NRC-1, utilizamos os aTSSs obtidos pelo experimento de dRNA-seq em conjunto com dados de bibliotecas de RNA-seq *paired-end* e fita específica disponíveis em nosso laboratório. Foi utilizado um total de 18 bibliotecas (Tabela 1). As bibliotecas foram processadas como descrito na seção 3.2. Os dados foram integrados e visualizados no Gaggles Genome Browser (BARE et al., 2010). Foi realizada uma inspeção visual buscando por sinais de transcritos definidos, ou seja, que comecem exatamente onde um aTSS foi mapeado, apresentando um declínio acentuado na cobertura de reads permitindo inferir sua extremidade 3' e observado em no mínimo quatro diferentes bibliotecas. Para auxiliar na inferência das extremidades 3' dos transcritos, foi utilizado o perfil de acúmulo de extremidades 5' de reads R2 obtido para cada biblioteca *paired-end*, considerando um mínimo de 10 reads R2 acumulados para a posição. Para os daTSSs, os mesmos critérios de anotação de um locus foram utilizados, com a restrição de que o sinal de transcrição deve apresentar no mínimo 10 nt de sobreposição com o gene *upstream* na fita oposta. O *pipeline* da anotação dos asRNAs está esquematizado na Figura 16.

Figura 16- Pipeline usado para anotação dos asRNAs.

3.6-Avaliação de sequências consenso na região promotora e extremidade 3'

A análise de conservação das regiões *upstream* e *downstream* aos aTSSs e daTSSs e também às extremidades 3' inferidas para os transcritos foram realizadas com a ferramenta online *Weblogo* (CROOKS et al., 2004) (disponível em <http://weblogo.berkeley.edu/logo.cgi>). Para a região promotora, as sequências extraídas para a análise correspondem a 40nt *upstream* a 10nt *downstream* a cada aTSS e daTSS. Para as extremidades 3', 10nt *upstream* e *downstream* ao final dos asRNAs anotados cuja extremidade 3' é correspondente a uma posição com no mínimo 10 reads R2 acumulados. Após obtenção das coordenadas genômicas com o mapeamento de aTSS e anotação dos transcritos, as sequências foram extraídas com a função *getfasta* do pacote *bedtools* v2.2.26 (QUINLAN, 2012) e para análise das extremidades 3' convertidas para sequências de RNA com a função *fasta_nucleotide_changer* do pacote *FASTX-Toolkit* (http://hannonlab.cshl.edu/fastx_toolkit/).

3.7-RIP-seq

Todo o experimento de RIPseq do qual os resultados foram utilizados nesse trabalho foram realizados pela Dra. Livia Soares Zaramela e serão resumidamente descritos abaixo para melhor compreensão dos dados que foram utilizados.

3.7.1-RIP-seq (RNA Immunoprecipitation – RNA-seq)

Para o ensaio de RIP-seq foram utilizadas duas linhagens de *H. salinarum* NRC1 em replicata biológica, uma superexpressando a proteína Lsm fusionada a uma cauda cmc (Lsm-cmc) e outra superexpressando apenas a cauda cmc (NRC1-cmc). As linhagens de *H. salinarum* NRC1-cmc (controle) e Lsm-cmc foram cultivadas até a densidade ótica (OD) de ~0,75. 20mL de suspensão celular foram centrifugadas, 3.700g, 10 minutos e ressuspensas em 12mL de solução basal (NaCl 250g/L; MgSO₄.7H₂O 20g/L; Citrato de Sódio 3g/L; KCl 2g/L). As suspensões celulares foram transferidas para placas de petri, no gelo, e irradiadas com luz no Crosslinker GE Healthcare UV 800x100 uJ/cm². A suspensão celular foi transferida cuidadosamente para tubos de 50mL e centrifugadas a 3.700g, 15 minutos, a 4°C. As células foram ressuspensas em 1mL de solução de lise (1X PBS; 0,1% SDS; 0,5% deoxicolato; 0,5% NP-40; inibidor de proteinase (1 tablete / 100mL – Sigma S8830), inibidor de RNase OUT Invitrogen 2µL/10mL) e deixadas no gelo por 5 minutos. A

suspensão foi centrifugada a 10.000g, 5 minutos, 4°C. O sobrenadante foi incubado com 10µL de *beads* (Invitrogen Dynabeads M-450 anti-mouse IgG, cat.#11041) por 10 minutos, a 4°C, para remover possíveis interações inespecíficas. A suspensão foi centrifugada a 10.000g, 5 minutos, 4°C. O sobrenadante foi incubado overnight, sob rotação, a 4°C, com 60µL de beads revestido com o anticorpo anti-cmyc (Sigma – M4439). Os beads foram lavados usando um suporte magnético apropriado. A lavagem foi realizada com 1mL de solução de lise, por duas vezes, 1mL de solução salina (5 X PBS, 0,1% SDS, 0,5% deoxicolato, 0,5% NP-40), por duas vezes, e com 1mL de TE (Tris-EDTA). Os *beads* foram ressuspensos em 100µL de TE e incubados a 65°C por 10 minutos. A suspensão foi centrifugada a 14.000g, 30 minutos, 25°C. Ao sobrenadante foi adicionado 120µL de TE/SDS (SDS 0,1%) e incubado por 30 minutos, a 65°C. Para isolar o RNA imunoprecipitado, a amostra foi incubada com 15µL de Proteinase K (Fermentas), por 2 horas, a 37°C. A purificação do RNA foi realizada por extração orgânica ou utilizando o kit MinElute Reaction Cleanup – Qiagen.

3.7.2 – RNA-seq de RNAs co-imunoprecipitados com a proteína LSm (RIP-Seq)

As amostras utilizadas na preparação das bibliotecas de RIP-seq foram quantificadas com o kit Quant-iT™ RiboGreen® RNA Assay (Invitrogen). Para preparação das bibliotecas utilizamos o kit TruSeq mRNA Stranded RNA Sample Preparation (Illumina) com uma quantidade inicial de 1µg de RNA. O protocolo foi executado conforme instrução do fabricante. Para assegurar que as bibliotecas fossem sequenciadas em proporções semelhantes foi realizada uma quantificação final por qPCR utilizando o kit KAPA Library Quant (KapaBiosystems), o qual possui *primers* específicos para os adaptadores ligados as moléculas. Após quantificação, o volume de cada biblioteca foi ajustado para que obtenhamos quantidades similares de sequências ao final do processo. As bibliotecas de RNAs imunoprecipitados com a proteína Lsm foram sequenciadas com o kit MiSeq Reagent v2 50 cycles single end(Illumina) no sequenciador MiSeq (Illumina).

3.7.3- Análise RIP-seq

As bibliotecas correspondentes as réplicas biológicas de cada linhagem foram colapsadas em único arquivo. A comparação entre as bibliotecas produzidas a partir de RNAs ligados à LSm e as controle foi realizada para inferir quais transcritos interagem com essa proteína. Os reads foram trimados utilizando Trimmomatic v0.36 e alinhados com HISAT2

v2.1.0. ao genoma de referência de *H. salinarum* NRC-1. Os arquivos SAM gerados no alinhamento foram convertidos para BAM utilizando SAMtools v1.3.1 (LI, HENG et al., 2009) e foram utilizados como entrada para o programa NMR para que a melhor posição de alinhamento dos reads mapeados em mais de uma posição fosse encontrada levando em conta a minimização da variância local da cobertura. As contagens de *reads* foram normalizadas em relação a maior biblioteca e o $\log_2$ do *fold change* (FC) da biblioteca de RNAs imunoprecitados com Lsm em relação a controle foi calculado para cada posição. Foram anotadas como regiões genômicas correspondentes a transcritos interagindo com LSm todas as regiões que possuem valor $\log_2(\text{FC}) \geq 1,5$ por no mínimo 10 posições consecutivas.

3.8-asRNAs interagindo com LSm

Para inferir quais transcritos antisense anotados interagem com a proteína LSm, foram encontradas as sobreposições entre os *loci* de *asRNAs* inferidos com as regiões genômicas correspondentes a transcritos interagindo com LSm inferidas como descrito na seção anterior, para isso foi utilizado a função *intersect* do pacote *bedtools* v2.2.26 (QUINLAN, 2012). As sobreposições foram filtradas utilizando *scripts* em R e apenas os *asRNAs* com no mínimo 10 nt de sobreposição com as regiões genômicas correspondentes a transcritos interagindo com LSm foram considerados interagindo com a proteína LSm.

3.9-Função dos genes sense

Para obter um panorama da função dos genes com *asRNAs*, utilizamos a ferramenta KEGG orthology (KO) para associação dos genes às suas possíveis funções (AOKI-KINOSHITA; KANEHISA, 2007). A lista de genes foi submetida para análise utilizando o webserver KEGG mapper (http://www.genome.jp/kegg/tool/map_brite1.html). Na ausência de um KO associado, dados provenientes do banco MicrobesOnline (<http://www.microbesonline.org/>) e a utilização da ferramenta BlastCD (MARCHLER-BAUER et al., 2017) para identificar domínios proteicos conservados, foram utilizados para associar funções.

3.10- Anotação de TAs tipo II e busca de aTSSs

Para a anotação dos TAs do tipo II no genoma de *E. coli* e das arqueias analisadas, utilizamos a plataforma online TA finder 2.0 (<http://202.120.12.133/Tafinder/index.php>) com

todos parâmetros em modo *default*. As posições dos aTSSs obtidas por dRNA-seq e as anotações dos asRNAs encontrados em TAs do tipo II em *E. coli* foram extraídas dos arquivos suplementares das publicações de Thomason et al. (2015) e Lybecker et al. (2014). Para mapear os TSS das demais arquéias, os dados dos experimentos de dRNA-seq foram baixados da base de dados SRA (*Haloferax volcanii* DS2 (PRJNA324298), *Methanocaldococcus jannaschii* DSM 2661 (PRJNA342613), *Thermococcus kodakarensis* KOD1 (PRJNA242777) e *Thermococcus onnurineus* NA1 (PRJNA339284)) e reanalisados conforma descrito na seção 3.3. O corte de p-valor para considerar um TSS foi o mesmo que para NRC-1, p-valor $\leq 0,005$.

3.11- Expressão diferencial

Para as análises de expressão diferencial um arquivo GFF foi gerado com os genes da anotação utilizada nesse estudo incluindo também os 846 asRNAs anotados, gerando um arquivo com um total de 3680 *features*. As matrizes de contagem de reads por *feature* foram geradas com o programa HT-Seq (ANDERS; PYL; HUBER, 2015) a partir dos arquivos *.bam* gerados após o alinhamento. As análises de expressão diferencial foram realizadas com o pacote DEseq2 (LOVE; HUBER; ANDERS, 2014), disponível na linguagem R. Foram considerados diferencialmente expressos os genes com valor absoluto de log2 do *fold change* maiores que 1 e menores que -1 e com p-valor ajustado menor que 0,01.

3.12- aTSSs conservados em *H. volcanii*

Para identificação de aTSSs possivelmente conservados entre *H. salinarum* NRC-1 e *H. volcanii*, os dados de dRNA-seq de *H. volcanii* foram baixados do SRA e reanalisados conforme descrito na seção 3.11. Buscamos por genes ortólogos compartilhados por essas duas arquéias utilizando a base de dados OrtholugeDB 2.1 (disponível em <http://www.pathogenomics.sfu.ca/ortholugedb/>) (WHITESIDE et al., 2013) e quais desses ortólogos possuem aTSSs. Os genes tiveram seus tamanhos normalizados e divididos em 10 regiões de mesmo tamanho. Consideramos como potencialmente conservados os aTSSs que pertencessem a mesma região nos correspondentes ortólogos de seus respectivos organismos.

3.13- Comparação entre sequências de RNases R

As sequências de aminoácidos das RNases R de *H. volcanii* e *H. salinarum* NRC-1 foram obtidas no banco MicrobesOnline (<http://www.microbesonline.org/>). Para verificar a existência de ambiguidade de genes que codificam essa RNase R no genoma de *H. salinarum* NRC-1, foi realizado um alinhamento local com Blastp (ALTSCHUL et al., 1990) utilizando a sequência de aminoácidos RNase R de *H. volcanii*. O alinhamento múltiplo das duas sequências foi realizado com o software Clustal Omega (SIEVERS et al., 2011) e a representação gráfica com a ferramenta ESPrpt 3.0 (ROBERT; GOUET, 2014). A busca por domínios conservados nas sequências foi realizada com BlastCD (MARCHLER-BAUER et al., 2017).

3.14 Construção da linhagem *ΔrnaseR*

As linhagens *H. salinarum* *ΔrnaseR* (NRC-1- Δ VNG_RS10370, Δ VNG_RS06485) e *Δura3* (NRC1- Δ VNG_RS06485) foram fornecidas pelo Laboratório do Professor Nitin Baliga (*Institute for Systems Biology – ISB*) (WURTMANN et al., 2014). As linhagens foram obtidas pelo método *pop-in-pop-out*, que se baseia no processo de recombinação homóloga entre o gene alvo no genoma e um fragmento de DNA contendo as regiões flangeadoras do gene alvo a ser deletado (KAUR et al., 2006; PECK et al., 2000). A linhagem *Δura3* é utilizada como linhagem precursora para nocautear outros genes de interesse. A ausência ou presença da proteína codificada por esse gene permite a seleção de linhagens recombinantes, etapa fundamental na construção de linhagens *knockout* em *H. salinarum* NRC-1 (PECK et al., 2000).

3.15- Curva de crescimento

Curvas de crescimento foram realizadas para avaliação fenotípica da linhagem *ΔrnaseR* comparada a linhagem controle *Δura3*. O experimento foi realizado em frascos Erlenmeyer de 125 ml contendo 40 ml de CM (250g/l NaCl, 20g/l MgSO₄, 2 g/l KCl, 3 g/l Citrato de Sódio, 10 g/l Peptona Bacteriológica Oxoid) com suplementação de uracila (50 µg/ml) sob agitação constante de 125 rpm, a 37°C e iluminação constante. As curvas foram construídas por leitura da densidade óptica (OD 600nm), lidas quatro vezes por dia durante cinco dias. A cada ponto de leitura, foram obtidas os valores de OD de triplicatas biológicas de cada linhagem. Para análise das curvas de crescimento, foram utilizadas funções do

programa R (R CORE TEAM, 2017). A média do valor de leitura obtido das triplicatas das linhagens por ponto foi plotado com os respectivos desvios padrões. Para confirmação do fenótipo observado, as curvas foram repetidas três vezes de forma independente.

3.16- Extração RNA total

Para extração de RNA total, as linhagens *ΔrnaseR* e *Δura3* foram crescidas em meio CM (250g/l NaCl, 20g/l MgSO₄, 2 g/l KCl, 3 g/l Citrato de Sódio, 10 g/l Peptona Bacteriológica Oxoid) complementado com uracila (50 μg/ml) e sob condições referência (225rpm, 37°C e iluminação constante) (DASSARMA et al., 1995). Quando a cultura atingiu a densidade óptica (OD 600nm) de cerca de 0,5 (metade da primeira fase *log* de crescimento), alíquotas de 2 ml de cultura foram coletadas em tubos eppendorfs e centrifugados a 14.000 rpm por 1 minuto. Os sobrenadantes foram descartados e os *pellets* obtidos foram armazenados a -80° C. O RNA total foi extraído dos *pellets* com o kit mirVana miRNA Isolation (Ambion) seguindo protocolo de extração de RNA total indicado pelo fabricante. Foram unidos dois *pellets* da mesma cultura para cada extração individual. Para cada amostra de RNA extraído, 50μg de RNA foram tratados com DNase para eliminação de DNA residual do processo de extração. Foram adicionados 10μl de Turbo DNase I (Fermentas), 15μl de Turbo DNaseI 10x buffer e 0,5 μl de Ribolock (Fermentas) para um volume final de 150μl para cada reação. As reações foram incubadas a 37 °C por 45 minutos. Ao término desse tratamento, as amostras foram precipitadas adicionando 150μl de fenol ácido, logo após vortexadas e centrifugadas a 10.000g por 5 minutos. A fase aquosa de cada amostra foi transferida para novo microtubo onde foram adicionados 2,5x o volume de Etanol 100% e 1/10 do volume de NaOAc 3M. As reações foram incubadas a -20°C *overnight*. As amostras foram centrifugadas a 14.000g, 4°C por 30 minutos, os sobrenadantes foram desprezados. Os *pellets* foram lavados com 500μl de etanol 70% e em seguida as amostras foram centrifugadas por 5 min a 10.000g. As amostras foram secas a 65°C por 10 minutos e eluídas em água MiliQ DEPC.

3.17 RNA-seq *ΔrnaseR*

O experimento foi realizado com duplicatas biológicas de cada linhagem. As amostras para o preparo das bibliotecas de RNA-seq foram quantificadas com o kit Quant-iT RiboGreen RNA Assay (Invitrogen). Para o preparo das bibliotecas utilizamos o kit TruSeq

Small RNA Sample Preparation (Illumina) seguindo protocolo para manter a informação de especificidade de fita das moléculas de RNA. Foi utilizada quantidade inicial de 1µg de RNA de cada amostra. Na primeira etapa do preparo das bibliotecas foi realizada a ligação de adaptadores as extremidades 3' e 5' das moléculas de RNA, em seguida, a transcrição reversa para converter as moléculas de RNA em cDNA seguida de uma amplificação por PCR que adiciona um *index* único a cada biblioteca. Essa amplificação também tem por finalidade aumentar o número de cópias das moléculas presentes na reação. As amostras amplificadas foram corridas em gel 6% Novex TBE PAGE (Invitrogen) e a porção entre os marcadores de 145nt e 500nt foi excisada para purificação, resultando em uma seleção de fragmentos com o tamanho máximo de aproximadamente 600nt, esse é o tamanho máximo em que as moléculas são capazes de formar ligações em *bridge* para obtenção dos *clusters* no *flowcell* e além disso, consiste em uma abordagem para reduzir a quantidade de rRNA nas amostras. Cada porção do gel correspondente as amostras foi extraída independentemente e os fragmentos foram eluídos em H2O MilliQ *overnight*. Ao final, as amostras eluídas foram purificadas em tubos com filtros de 5 µm. Para assegurar que as bibliotecas fossem sequenciadas em proporções semelhantes, foi realizada uma quantificação final por qPCR utilizando o kit KAPA Library Quant (Kapa Biosystems), o qual possui *primers* específicos para amplificação das moléculas ligadas aos adaptadores. As bibliotecas foram sequenciadas com o kit MiSeq Reagent v2 300 cycles *paired-end* (Illumina) no sequenciador MiSeq. Os dados foram processados e analisados conforme descrito na sessão 3.2 e a análise de expressão diferencial conforme descrito na sessão 3.12.

4-Resultados

4.1-Mapeamento da extremidade 5' de asRNAs: aTSSs

Utilizamos a anotação disponível no Refseq da linhagem NRC-1 com a anotação curada manualmente da linhagem R1, resultando em 2834 genes considerados para o mapeamento de asRNAs. Foram mapeados 2146 aTSS em 1231 genes (Apêndice A – Materiais suplementares - S1). Eventuais estruturas secundárias formadas pelos RNAs podem prevenir a degradação dos mesmos pela enzima TEX, de modo que, transcritos contendo tais estruturas podem ser enriquecidos na amostra tratada com TEX não por serem primários, mas por conterem estruturas não degradadas pela enzima levando ao mapeamento de falsos TSSs (JÄGER et al., 2014). Esses possíveis falsos positivos foram filtrados utilizando os valores MFE (*minimum free energy*) de estruturas preditas em regiões *downstream* as posições de cada aTSS (Materiais e métodos, seção 3.4). Após aplicação do filtro de estruturação de RNAs, restaram 1626 aTSS distribuídos em 963 genes (Apêndice A – Materiais suplementares - S2).

A Figura 17A mostra a posição relativa dos aTSS mapeados em relação ao tamanho de seus genes cognatos normalizados em porcentagem. Pode-se notar um acúmulo de aTSSs nas extremidades dos genes, sendo que a maior parte está localizada na extremidade 3'.

Figura 17- Posições relativas dos aTSSs. Os genes com aTSSs mapeados tiveram seu tamanho normalizado e foram divididos em 10 regiões. **A** Posição relativa de todos 1626 aTSS. **B** Posição relativa dos 846 aTSSs que possuem um asRNA anotado.

4.2-Anotação dos asRNAs: inferência das extremidades 3'

Para a anotação dos RNAs antisense associados aos aTSS encontrados, foram utilizadas 18 bibliotecas de RNA-seq fitas específicas e *paired-end*. Após processamento e alinhamento ao genoma de referência, foi obtido um total de 22.534.041 pares R1-R2 alinhados, 2.443.967 apenas R1 alinhados e 1.646.985 apenas R2 alinhados. As quantidades de reads sequenciados, trimados e alinhados para cada biblioteca podem ser observadas na Tabela 1. Os dados foram plotados para visualização e curagem manual dos asRNA no software Gaggie Genome browser (Figura 18).

Figura 18- Screenshot GGB. Cobertura de reads (log2 da contagem de reads por posição) de todas as 18 bibliotecas PE para uma região do genoma de *H. salinarum* NRC-1. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. Triângulos visualizadas no GGB. pretos indicam TSSs mapeados por dRNA-seq.

Para auxiliar na anotação das extremidades 3' dos transcritos, utilizamos um perfil de acúmulo de extremidades 5' dos reads R2 ao longo do genoma de referência para guiar a anotação manual. Dos 1626 aTSS, definimos 766 asRNA *loci* distribuídos em 558 genes (Apêndice A – Materiais suplementares - S3), para esses, a extremidade 3' do transcrito foi anotada exatamente onde havia um acúmulo de extremidades 5' de reads R2 acompanhado de uma redução acentuada no sinal de cobertura por reads em 425 casos (Figura 19), para os demais a extremidade foi inferida tomando como referência apenas o decaimento acentuado no sinal do transcrito. 26 dos 766 asRNAs *loci* estão antisense a dois genes concomitantemente (Figura 20). A maioria dos genes apresenta apenas um asRNA *loci*

Tabela 1 – Número de reads sequenciados, trimados e alinhados para cada biblioteca.

Biblioteca	Preparo	Pares de reads sequenciados	Trimados como par	Trimados apenas R1	Trimados apenas R2	Pares alinhados	R1 alinhados	R2 alinhados
ΔRNase R -1	RNA total	3280188	1695364 (51.68%)	365360 (11.14%)	124600 (3.80%)	1609712 (94.95%)	347117 (95.01%)	116610 (93.59%)
ΔRNase R -2	RNA total	4846385	2060439 (42.51%)	525883 (10.85%)	142498 (2.94%)	1947464 (94.52%)	497454 (94.59%)	130706 (91.72%)
ΔuraA -1	RNA total	2893173	1304294 (45.08%)	327468 (11.32%)	87443 (3.02%)	1232193 (94.47%)	312215 (95.34%)	81653 (93.38%)
ΔuraA -2	RNA total	2375779	1081316 (45.51%)	266161 (11.20%)	69093 (2.91%)	1008636 (93.28%)	250776 (94.22%)	63905 (92.49%)
REF -1	RNA total	1878153	728024 (38,76%)	90608 (4,82%)	76221 (4,06%)	639768 (87.88%)	79467 (87.70%)	68550 (89.94%)
REF -2	RNA total	2378435	937497 (39,42%)	115541 (4,86%)	109256 (4,59%)	836256 (89.20%)	101459 (87.81%)	99080 (90.69%)
REF -3	RNA total	1245709	678514 (54,47%)	36052 (2,89%)	12985 (1,04%)	644835 (95.04%)	31559 (87.54%)	11385 (87.68%)
REF -4	RNA total	945496	698132 (73,84%)	43424 (4,59%)	11493 (1,22%)	654989 (93.82%)	39860 (91.79%)	10342 (89.99%)
17h -1	RNA total	2175004	1529197 (70,31%)	175497 (8,07%)	49708 (2,29%)	1386400 (90.66%)	164029 (93.47%)	46829 (94.21%)
17H -2	RNA total	1891117	1191182 (62,99%)	129348 (6,84%)	31557 (1,67%)	1076385 (90.36%)	116862 (90.35%)	29519 (93.54%)
37h -1	RNA total	3238764	2396037 (73,98%)	199282 (6,15%)	75965 (2,35%)	2227579 (92.97%)	185893 (93.28%)	70579 (92.91%)
37h -2	RNA total	2264181	1753053 (77,43%)	171178 (7,56%)	49424 (2,18%)	1626322 (92.77%)	162604 (94.99%)	46207 (93.49%)
86h -1	RNA total	1532836	800314 (52,21%)	55119 (3,60%)	16561 (1,08%)	756981 (94.59%)	47326 (85.86%)	14441 (87.20%)
86h -2	RNA total	1570241	889164 (56,63%)	47459 (3,02%)	18003 (1,15%)	840445 (94.52%)	41379 (87.19%)	15702 (87.22%)
small-rna I -1	Pequenos RNAs	2616357	1354007 (51.75%)	109457 (4.18%)	269630 (10.31%)	1197695 (88,46%)	91984 (84.04%)	242982 (90,11%)
small-rna I -2	Pequenos RNAs	2089112	1000408 (47.89%)	79196 (3.79%)	199649 (9.56%)	889692 (88.93%)	65967 (83.30%)	180014 (90.17%)
small-rna II -1	Pequenos RNAs	5464209	2683630 (49.11%)	732141 (13.40%)	451837 (8.27%)	2411273 (89.85%)	659579 (90.09%)	418481 (92.62%)
small-rna II -2	Pequenos RNAs	4630249	1685715 (36.41%)	427200 (9.23%)	269888 (5.83%)	1547416 (91.80%)	384166 (89.93%)	250291 (92.74%)
RIP-sep control*	RNAs imunoprecipitados	6369053		5903958 (92.70%)			3145380 (53.28%)	
RIP-seq trat*	RNAs imunoprecipitados	12941962		12158948 (93.95%)			6423287 (52.83%)	

* Bibliotecas sequenciadas em modo single-end.

associado, porém foram encontrados casos de genes com até 5 asRNAs *loci* anotados (Figura 26). A Figura 17B mostra a posição relativa dos aTSS com *locus* em relação aos seus genes cognatos, pode ser observado um acúmulo nos extremos dos genes mas de modo geral, essa distribuição indica que um transcrito antisense pode começar de qualquer região. Para os 860 aTSS restantes, não foi detectado sinal de expressão gênica definido ou não foi possível inferir a extremidade 3' do transcrito associado ao aTSS.

Figura 19- Dados de bibliotecas PE para anotação de asRNAs. Exemplo de um asRNA (seta rosa) anotado para o gene VNG_RS01595 (seta laranja) que codifica uma peptidase. R1 e R2 se referem as bibliotecas obtidas em um sequenciamento *paired-end*, SE: *single-end*, PE: *paired-end*. Triângulo preto mostra um aTSS. Triângulo cinza mostra a posição da extremidade 5' dos *reads* da biblioteca R2. A primeira linha em vermelho mostra o sinal de RNA-seq (log2 da contagem de reads por posição) apenas de bibliotecas R1 SE para a fita *forward*. A segunda linha em roxo mostra o sinal de RNA-seq apenas de bibliotecas R2, SE. A terceira linha mostra o sinal de RNA-seq gerado após processamento em modo PE das bibliotecas R1 e R2. Esse exemplo ilustra como o uso de dados PE podem aumentar a precisão da inferência de extremidades 3' de um transcrito.

Figura 20- asRNA sobrepondo dois genes. Setas laranjas representam dois genes codificadores de proteínas ribossomais da subunidade 30S em um *operon*. Sinal de RNA-seq (log2 da contagem de reads por posição) para fita *forward* no painel superior e para a fita *reverse* no painel inferior. Triângulos pretos indicam TSSs. O *locus* de asRNA (seta rosa) definido está localizado na fita *forward*, possui 168 nt e sobrepõe os dois genes desse *operon*.

Realizamos uma busca por *downstream* aTSS (daTSS): TSS localizados na fita oposta a genes anotados, localizados até 200 nt downstream a extremidade 3' do gene e com sinal de expressão sobreposto ao gene upstream. Estes critérios foram utilizados baseados no fato de (i) asRNAs podem iniciar em regiões *downstream* ao gene cognato, (ii) regiões 3'UTR de mRNAs podem ser alvos para o pareamento de ncRNAs e (iii) asRNAs oriundos das regiões 3'UTR podem representar um importante grupo de ncRNAs reguladores em arqueias sem 5'UTRs longos (BABSKI et al., 2014; JÄGER et al., 2014); LYBECKER et al., 2014). Encontramos 80 daTSSs que dão origem a asRNAs (Apêndice A – Materiais suplementares - S3). Para 51 desses casos, a extremidade 3' do transcrito foi anotada exatamente onde havia um acúmulo de extremidades 5' de reads R2. Apenas um dos asRNAs anotados a partir de um daTSS sobrepõe dois genes (Figura 28).

No total, definimos 846 loci de asRNAs distribuídos em 613 genes (Apêndice A – Materiais suplementares - S3). A distribuição de tamanho dos 846 asRNA *loci* (766 asRNAs + 80 daRNAs) é mostrada na (Figura 21), mostrando que a maioria é constituída de RNAs pequenos, menores do que 100 nt.

Figura 21- Distribuição de tamanho dos 846 asRNAs loci definidos a partir dos aTSSs.

Para verificar se a nossa abordagem de inferência de 3' dos transcritos associados aos aTSSs revela algum tipo de padrão de sequência no término desses transcritos, extraímos sequências de 10 nt *upstream* e *downstream* ao final dos asRNAs anotados. Essa análise foi feita somente para os asRNAs cuja extremidade 3' é correspondente a uma posição com mais de 10 reads R2 acumulados, buscando padrões de sequências associadas ao término dos asRNA inferidos (Figura 22). Fica claro um enriquecimento de purinas, principalmente uracila na última (0) e penúltima (-1) base dos transcritos.

Figura 22- Frequência nucleotídica das extremidades 3' dos asRNAs anotados. Posição 0 indica a última base do transcrito.

4.3- asRNAs sobrepostos a extremidade 5' dos genes

Bouvier et al. (2008) mostraram experimentalmente que não é necessário que um sRNA bloqueie os sítios de reconhecimento da maquinaria ribossomal localizados na região 5'UTR para que a inibição da tradução por pareamento antisense ocorra. No sRNA analisado no estudo, pareamentos que ocorrem até o quinto códon localizado no mRNA são o bastante para impedir a ligação da maquinaria ribossomal e inibir a tradução (BOUVIER et al., 2008). Tomando por base essa informação, buscamos no nosso conjunto de asRNAs anotados aqueles que sobrepõem as primeiras bases correspondentes ao códon de início de tradução de seus genes cognatos ou aqueles cuja extremidade 3' estão localizados a no máximo 12 nt da extremidade 5' da CDS, ou seja, aqueles que mesmo sem sobrepor a extremidade 5' ainda sobrepõem no mínimo as bases correspondentes ao quinto códon do mRNA de seu gene.

cognato. Dos 846 loci definidos, encontramos 145 asRNAs com essas características (Figura 23) (Apêndice A – Materiais suplementares - S4).

Figura 23- Distribuição da distância em nt da extremidade 3' dos asRNAs anotados até a extremidade 5' do gene cognato. Valores menores que zero indicam asRNAs que terminam após a extremidade 5' do gene cognato.

4.4-Expressão em relação ao gene sense

Levando em conta a média da contagem de reads normalizada das réplicas biológicas das bibliotecas de RNA-seq dos experimentos utilizados nesse estudo, observamos que a maioria dos asRNAs definidos possuem expressão baixa em relação a seus genes sense. Apenas 118 asRNA (Apêndice A – Materiais suplementares - S5) identificados apresentam expressão média maior ou igual à de seu gene cognato em, ao menos, um dos pares de réplicas biológicas analisadas nesse trabalho. Na Figura 24, mostramos como exemplo a dispersão e a relação do $\log_2$ do *fold-change* da cobertura média da contagem normalizada de *reads* dos asRNAs em relação a de seus mRNAs para quatro de nossas bibliotecas, as demais apresentam o mesmo perfil de cobertura.

Figura 24- Cobertura de *reads* dos asRNAs em relação a seus genes cognatos. Gráfico de dispersão mostra a cobertura de *reads* dos genes no eixo X e dos asRNAs anotados no eixo Y. O histograma mostra a distribuição de valores do $\log_2$ do *fold change* da cobertura do asRNA em relação ao gene

cognato. Média da cobertura normalizada das réplicas das bibliotecas: Δ RNase R (A-B), 17h(C-D), 86h(E-F), small-rna I (G-H).

4.5-Região Promotora

Analizamos a frequência dos nucleotídeos da região 40 nt *upstream* e 10 nt *downstream* em relação aos aTSSs e daTSSs (Figura 25). Pode ser notada a predominância de purinas na posição 0, correspondente a posição do aTSS e de pirimidinas na posição -1. Também é possível observar assinaturas de sequências correspondentes a elementos característicos de regiões promotoras em arquéias, sendo esses uma TATA-box na posição -26, e o sítio de reconhecimento do fator de transcrição TIIB (B recognition element – BRE) caracterizado pelo enriquecimento de nucleotídeos GC na posição -36 (BELL; JACKSON, 2001; SEITZER et al., 2012).

Figura 25- Análise da frequência nucleotídica nas regiões promotoras e de início de transcrição dos asRNAs anotados.

4.6- asRNAs que interagem com LSm

Identificamos 91 asRNAs com região de interação com LSm (Apêndice A – Materiais suplementares - S6), desses o transcrito do gene sense também interage com LSm em 81 casos, podendo constituir complexos asRNA-Lsm-mRNA. 48 desses asRNAs pertencem a genes que codificam proteínas hipotéticas. 7 dos asRNAs interagindo com LSm pertencem a genes que codificam elementos transponíveis.

Dois exemplos de genes com asRNAs anotados e regiões de interação com LSM em transcritos correspondentes as fitas *forward* e *reverse* são os genes VNG_RS00195 e VNG_RS13220 (Figura 26), ambos codificam glicosiltransferases da família 1. Esses genes estão provavelmente envolvidos na formação da *S-layer*, uma camada de glicoproteínas que envolve a membrana plasmática de *H. salinarum* NRC-1 (DYALL-SMITH, 2008).

Figura 26- asRNAs *loci* anotados para os genes VNG_RS00195 e VNG_RS13220. (setas laranjas, anotados na fita *reverse*), que codificam glicosiltransferases do tipo 1. Setas rosas indicam os *loci* de asRNAs anotados. Regiões de interação com LSM estão indicadas com linhas verdes, TSSs representados por triângulos pretos. Em vermelho o perfil do sinal dos dados de RNA-seq (log2 da contagem de reads por posição).

4.7-Função dos genes sense potencialmente regulados por asRNAs

Do total de 613 genes que possuem ao menos um asRNA *locus*, 198 estão anotados como proteínas hipotéticas e 60 como elementos transponíveis (Apêndice A – Materiais suplementares - S3). Para os demais 355 genes anotados, utilizamos a ferramenta KEGG orthology (KO) para associação dos genes às suas possíveis funções (AOKI-KINOSHITA; KANEHISA, 2007) (Figura 27).

Figura 27- Representatividade dos KOs identificados nos genes com asRNAs anotados.

4.7.1-asRNAs em genes relacionados a Tradução

Anotamos asRNAs para 37 genes relacionados a tradução (Tabela 2), incluindo genes que codificam proteínas ribossomais, fatores de iniciação de tradução e tRNAs Asn, Lys e Ser e tRNAs ligases

4.7.2-asRNAs em genes codificadores de Transposases

105 dos 846 asRNAs anotados estão sobrepostos a genes que codificam transposases em *H. salinarum* NRC-1 (Apêndice A – Materiais suplementares - S3). Em bactérias, exemplos de asRNAs inibindo a tradução do mRNA de transposases inibindo a ligação da maquinaria ribossomal a extremidade 5' são conhecidos (ELLIS; TRUSSLER; HANIFORD, 2015). Encontramos 16 asRNAs sobrepostos a extremidade 5' de genes codificadores de transposases (Tabela 3).

Tabela 2 – asRNAs em genes relacionados a tradução.

replicon	cod. ATSS	fita	inicio	fim	tamanho(nt)	gene	anotacao
chr	aTSS_111	+	84173	84248	75	VNG_RS00390	VNG0099G 50S ribosomal protein L16
chr	aTSS_207	+	197207	197257	50	VNG_RS00970	VNG0239C 50S ribosomal protein L37Ac
chr	aTSS_653	+	845383	845406	23	VNG_RS04305	VNG1103G 50S ribosomal protein P1
chr	aTSS_654	+	847089	847176	87	VNG_RS04315	VNG1105G 50S ribosomal protein L1
chr	aTSS_660	-	860587	860616	29	VNG_RS04410	VNG1132G 30S ribosomal protein S13
chr	aTSS_661	-	860768	860876	108	VNG_RS04415	VNG1133G 30S ribosomal protein S4
chr	aTSS_662	-	863354	863476	122	VNG_RS04445	VNG1139G,VNG1139Gm 30S ribosomal protein S9
chr	aTSS_933	-	1260898	1260951	53	VNG_RS06565	VNG1691G 50S ribosomal protein L23
chr	aTSS_937	-	1267447	1267527	80	VNG_RS06635	VNG1709G 50S ribosomal protein L6
chr	aTSS_939	-	1268699	1268722	23	VNG_RS06645	VNG1713G 50S ribosomal protein L19e
chr	aTSS_734	+	958849	958886	37	VNG_RS04995	VNG1276C ribosome assembly factor SBDS
chr	aTSS_735	+	959429	959465	36	VNG_RS04995	VNG1276C ribosome assembly factor SBDS
chr	daTSS_60	+	1952651	1952727	76	VNG_RS10220	VNG2612G ribosome biogenesis/translation initiation ATPase RLI
chr	aTSS_965	-	1296499	1296526	27	VNG_RS06805	VNG1756G translation initiation factor 1A 2
chr	daTSS_60	+	1952651	1952727	76	VNG_RS10220	VNG2612G ribosome biogenesis/translation initiation ATPase RLI
chr	aTSS_149	+	114689	114905	216	VNG_RS00560	VNG0137G tRNA CCA-pyrophosphorylase
chr	daTSS_12	-	115158	115258	100	VNG_RS00565	VNGt04 tRNA-Asn
chr	daTSS_18	-	311134	311319	185	VNG_RS01565	VNG0403G proline--tRNA ligase
chr	daTSS_26	-	748766	748896	130	VNG_RS03830	VNGt14 tRNA-Lys
chr	daTSS_30	-	860107	860265	158	VNG_RS04405	VNGt17 tRNA-Ser
chr	aTSS_668	-	872269	872436	167	VNG_RS04500	VNG1153G glutamate--tRNA ligase
chr	aTSS_669	-	872953	872994	41	VNG_RS04500	VNG1153G glutamate--tRNA ligase
chr	aTSS_838	+	1122703	1122803	100	VNG_RS05875	G1510C tRNA uridine(34) 5-carboxymethylaminomethyl modification radical SAM/GNAT enzyme E
chr	daTSS_46	+	1313380	1313492	112	VNG_RS06880	VNG1774G glutamyl-tRNA reductase
chr	aTSS_1007	+	1355915	1356041	126	VNG_RS07110	VNG1835G threonine--tRNA ligase
chr	aTSS_1115	-	1525471	1525501	30	VNG_RS08010	VNG2072G serine--tRNA ligase
chr	aTSS_1123	-	1535292	1535479	187	VNG_RS08085	VNG2088G tRNA (guanine(10)-N(2))-dimethyltransferase
chr	aTSS_1191	-	1621346	1621451	105	VNG_RS08540	VNG2190G isoleucine--tRNA ligase
chr	aTSS_1205	+	1638843	1638886	43	VNG_RS08610	VNG2208G tryptophan--tRNA ligase
chr	aTSS_1206	+	1638888	1639052	164	VNG_RS08610	VNG2208G tryptophan--tRNA ligase
chr	aTSS_1217	+	1654026	1654062	36	VNG_RS08665	VNG2223G leucine--tRNA ligase
chr	aTSS_1223	+	1661793	1661856	63	VNG_RS08695	VNG2232G tryptophan--tRNA ligase
chr	aTSS_1255	-	1699116	1699230	114	VNG_RS08910	VNG2283G alanine--tRNA ligase
chr	aTSS_1257	-	1699801	1699846	45	VNG_RS08910	VNG2283G alanine--tRNA ligase
chr	aTSS_1380	-	1870901	1871009	108	VNG_RS09770	VNG2504G phenylalanine--tRNA ligase subunit alpha
chr	aTSS_1414	+	1909933	1910090	157	VNG_RS09965	VNG2547G valine--tRNA ligase
plasmid_pNRC200	aTSS_2023	-	241885	242155	270	VNG_RS11615	VNG6312G arginine--tRNA ligase

Tabela 3 – asRNAs sobrepostos a extremidade 5' de genes codificadores de transposases.

Replicon	cod. ATSS	fit	inicio_asRNA	fim_asRNA	tamanho(nt)	gene	anotacao	inicio_gene	fim_gene	Sobreposicao_5'
chr	aTSS_224	-	229047	229078	31	VNG_RS01140VNG0285C	IS200/IS605 family transposase ISH22	229052	229441	-5
chr	aTSS_883	+	1189718	1189773	55	VNG_RS06190	partial;pseudo;VNG1590H transposase	1189491	1189760	-13
chr	aTSS_1488	+	1989174	1989204	30	VNG_RS10400	VNG2652H transposase	1987884	1989188	-16
plasmid_pNRC100	aTSS_1533	+	3715	3762	47	VNG_RS12220	VNG7005 ISNCY family transposase ISH7A	1905	3719	-43
plasmid_pNRC100	aTSS_1583	-	32084	32113	29	VNG_RS12360	ISH3 family transposase ISH3C	32090	33256	-6
plasmid_pNRC100	aTSS_1591	-	34515	34554	39	VNG_RS12370	VNG7034 IS5 family transposase ISH28	34534	35361	-19
plasmid_pNRC100	aTSS_1598	-	36810	36837	27	VNG_RS12385	VNG7036 IS5 family transposase ISH11	36819	37811	-9
plasmid_pNRC100	aTSS_1697	-	101483	101511	28	VNG_RS12710	ISH3-like element ISH27-3 family transposase	101488	102654	-5
plasmid_pNRC100	aTSS_1806	-	150918	150946	28	VNG_RS12925	ISH3 family transposase ISH27-2	150923	152089	-5
plasmid_pNRC100	aTSS_1861	+	186552	186576	24	VNG_RS13110	VNG7173 IS5 family transposase ISH11	185578	186570	-6
plasmid_pNRC200	aTSS_1891	+	125062	125096	34	VNG_RS11075	VNG6148H IS4 family transposase ISH32	123755	125089	-7
plasmid_pNRC200	aTSS_1904	+	129125	129151	26	VNG_RS11085	ISH3 family transposase ISH27-2	127982	129148	-3
plasmid_pNRC200	aTSS_1910	+	134804	134831	27	VNG_RS11105	ISH3 family transposase ISH3C	133661	134827	-4
plasmid_pNRC200	aTSS_2052	-	261266	261295	29	VNG_RS11685	ISH3 family transposase ISH3C	261272	262438	-6
plasmid_pNRC200	aTSS_2074	+	274277	274304	27	VNG_RS11735	ISH3 family transposase ISH3C	273134	274300	-4
plasmid_pNRC200	aTSS_2106	+	295462	295510	48	VNG_RS11820	partial;pseudo ISH3 family transposase	294596	295494	-16

*Sobreposicao_5': Distância em nt da extremidade 3' do asRNA até a extremidade 5' do gene sense. Valores negativos indicam lue o fim do asRNA se encontra upstream ao inicio do gene.

4.7.3-asRNAs nos genes codificadores de Rodopsinas

Identificamos um asRNA presente no gene *bop* (VNG_RS05715) que codifica a proteína de membrana bacteriorodopsina. Além desse, também identificamos outro asRNAs provavelmente associado a regulação do gene *bop*. Esse segundo asRNA identificado, se encontra antisense aos genes *brz* (VNG_RS05710), que codifica Brz (bacteriorodopsin regulating zinc finger protein) e *brb* (OE3105F), que codifica Brb (*bacteriorhodopsin regulating basic protein*). Estes dois genes estão localizados *upstream* ao gene *bop*. O asRNAs se inicia *downstream* ao gene *brb*, sobrepondo-o completamente e estendendo-se até aproximadamente metade do gene *brz* (Figura 28).

Figura 28- asRNA anotado para o gene *brb*. No painel superior, em vermelho, é mostrado o perfil do sinal dos dados de RNA-seq ($\log_2$ da contagem de reads por posição) para fita *forward* e no painel inferior para a fita *reverse*. Os genes *brz* (VNG_RS05710) e *brb* (OE3105F) estão na fita forward e estão representados em setas amarelas. Triângulos pretos indicam TSSs. O asRNA locus (seta rosa) foi definido a partir de um daTSS.

Também identificamos um asRNA para o gene VNG_RS00745 que codifica para a proteína halorodopsina. Essa proteína possui um grupo retinal que sofre modificações quando interage com a luz e atua como uma bomba de íons pelos mesmos princípios que a bacteriorodopsina, porém se distingue dessa última por bombear íons de cloreto através da membrana (OESTERHELT; TITTOR, 1989). A Tabela 4 contém as coordenadas dos três asRNAs encontrados nos genes relacionados as rodopsinas.

Tabela 4 – asRNAs nos genes codificadores e reguladores de rodopsinas.

replicon	cod. ATSS	fit	início	fim	tamanho(nt)	gene	anotacao
chr	aTSS_175	+	155806	155906	100	VNG_RS00745	VNG0180G halorhodopsin
chr	daTSS_36	-	1088797	1089100	303	VNG_OE3105Frb (OE3105F)	bacteriorhodopsin-regulating basic protein
chr	aTSS_824	-	1089545	1089615	70	VNG_RS05715	VNG1467G bacteriorhodopsin

4.7.4-asRNAs em genes relacionados a vesículas de gás – (*gas vesicle proteins - gvp*)

Identificamos três asRNAs para o gene *gvpD* da *vac region gvp-1* de *H. salinarum* NRC-1 (Figura 29), um deles sobreposto a extremidade 5'. Obtivemos a sequência da sonda utilizada por Kruger e Pfeifer (1996) para verificar a posição do asRNA da extremidade 5' previamente detectado por esses autores. Utilizando a ferramenta Blastn, verificamos sua posição no plasmídeo pNRC100. Quando comparamos a posição da sonda utilizada, “22131-22229”, com a posição do aTSS correspondente ao asRNA que anotamos, “22128”, verificamos que o último se encontra a apenas 3 nt *downstream* ao início da sonda utilizada (Figura 29). Uma das bandas obtidas como resultado do *Northern Blot* do estudo de 1996 é uma banda de 190 nt, nosso *locus* anotado possui tamanho próximo, de 164nt. Também identificamos asRNAs para os genes, *gvpCHIJL* (Tabela 5). Além desses, apesar de não conter um TSS, um nítido sinal de RNA-seq, provavelmente correspondente a um asRNA(pNRC100, “22447, 22509”, 63 nt), pode ser encontrado sobrepondo a extremidade 5' do gene *gvpA* (Figura 29).

Figura 29- asRNAs presentes nos genes *gvpD*, *gvpA* e *gvpC* da *vac region gvp1*. O gene *gvpD* (seta laranja) está codificado na fita reverse, os genes *gvpA* e *gvpC* na fita *forward* (setas amarelas). A região sombreada em azul correspondente a posição da sonda utilizada para o northernblot (KRUGER; PFEIFER, 1996). Em vermelho o perfil do sinal dos dados de RNA-seq (log2 da contagem de reads por posição). Triângulos pretos representam TSSs.

Tabela 5 – asRNAs nos genes *gvp* da região *vac gvp-1*.

replicon	cod. ATSS	fit	inicio	fim	tamanho(nt)	gene	anotacao
plasmid_pNRC100	aTSS_1555	+	16743	16863	120	VNG_RS12280	VNG7016 protein gvpL
plasmid_pNRC100	aTSS_1556	+	17055	17138	83	VNG_RS12280	VNG7016 protein gvpL
plasmid_pNRC100	aTSS_1557	+	18092	18168	76	VNG_RS12290	VNG7018 gas vesicle protein GvpJ
plasmid_pNRC100	aTSS_1558	+	18390	18428	38	VNG_RS13760	protein gvpI
plasmid_pNRC100	aTSS_1559	+	18615	18643	28	VNG_RS13760	protein gvpI
plasmid_pNRC100	aTSS_1560	+	18698	18722	24	VNG_RS12295	VNG7020 protein gvpH 1
plasmid_pNRC100	aTSS_1565	+	21000	21165	165	VNG_RS12315	VNG7024 protein gvpD
plasmid_pNRC100	aTSS_1567	+	22084	22106	22	VNG_RS12315	VNG7024 protein gvpD
plasmid_pNRC100	aTSS_1568	+	22128	22292	164	VNG_RS12315	VNG7024 protein gvpD
plasmid_pNRC100	aTSS_1569	-	22865	22964	99	VNG_RS12325	VNG7026 gas vesicle protein GvpC
plasmid_pNRC100	aTSS_1570	-	23888	23940	52	VNG_RS12325	VNG7026 gas vesicle protein GvpC

4.7.5-asRNAs em TAs tipo II em bactérias e arqueias

Identificamos asRNAs para os genes VNG_RS11240 e VNG_RS00140, ambos anotados como antitoxinas em diferentes TAs do tipo II (Apêndice A – Materiais suplementares - S3). Utilizando os dados de dRNA-seq para *E. coli* do estudo de Thomason et al. (2015), verificamos se com dados gerados em experimentos utilizando essa técnica podemos detectar os mesmos asRNAs reportados por Lybecker et al. (2014). Encontramos aTSSs para os dois genes codificadores de toxinas em que Lybecker et al. (2014) encontraram asRNAs. Para um desses genes codificadores de toxina, *yoeB*, que codifica uma endonuclease, integrante do TAs do tipo II *yoeB/yefM*, o aTSS mapeado por Thomason et al. (2015) se encontra a apenas 2 nt *downstream* ao início do asRNA identificado por Lybecker et al. (2014) (Figura 30).

Figura 30- asRNA e aTSS identificados para o gene *yoeB* que codificada uma toxina em um TAs do tipo II em *E. coli*. O asRNA identificado por Lybecker et al. (2014) possui as coordenadas “2087410-2087484” e o aTSS mapeado por Thomason et al. (2015) “2087412”, ambos na fita *forward* no genoma de *E. coli*.

Esses resultados nos levaram a hipótese de que a presença de asRNAs em TAs do tipo II pode se tratar de um fenômeno conservado e detectável por dRNA-seq em arqueias e bactérias. Para testar tal hipótese, utilizamos o programa *TA finder* 2.0 (<http://202.120.12.133/TAfinder/index.php>) para fazer uma anotação precisa de TAs do tipo II nos genomas de arqueias que possuem dados de dRNA-seq. Anotamos TAs do tipo II para os genomas de *H. salinarum* NRC-1, *H. volcanii* DS2, *M. jannaschii* DSM 2661, *T. kodakarensis* KOD1, *T. onnurineus* NA1 e também para o genoma de *E. coli*. A Tabela 6 mostra quantos TAs do tipo II foram anotados para cada organismo e quantos genes estavam previamente anotados.

Tabela 6: Número de TAs do tipo II previamente anotados e anotados nesse estudo com a ferramenta *TA finder*.

Organismo	Previamente anotados			TA finder
	Pares completos	Toxinas	Antitoxina	Pares identificados
<i>H. salinarum</i> NRC-1	1	1	2	11
<i>H. volcanii</i> DS2	0	5	0	9
<i>M. jannaschii</i> DSM 2661	2	5	0	15
<i>T. kodakarensis</i> KOD1	1	18	0	52
<i>T. onnurineus</i> NA1	0	1	0	8
<i>E. coli</i> str. K-12 substr. MG165 [†]	12	3	1	17

Utilizando as posições dos genes das toxinas e antitoxinas anotados, buscamos por TSSs mapeados na fita oposta desses genes (Apêndice A – Materiais suplementares – S7, S8, S9, S10, S11). Na Tabela 7, mostramos o número de toxinas e antitoxinas contendo ao menos um aTSSs nas arqueias com TAs do tipo II anotados e em *E. coli*. A maioria dos pares TAs anotados apresentam ao menos um aTSS em, pelo menos, um dos constituintes do par. Apenas em *H. salinarum* NRC-1 e *H. volcanii*, o número de toxinas com aTSSs não é maior que o de antitoxinas. Para a arqueia *T. onnurineus*, não foram encontrados aTSSs em nenhuma toxina ou antitoxina.

Tabela 7: TAs do tipo II com aTSSs mapeados.

Organismo	Gene	Com aTSS	*Com asRNA	aTSS em T ou A	aTSS em T e A	asRNA em ambos T/A
<i>H. salinarum</i> NRC-1	T	4	3	9	4	2
	AT	7	5			
<i>H. volcanii</i> DS2	T	4	-	6	2	-
	AT	4	-			
<i>M. jannaschii</i> DSM 2661	T	6	-	8	-	-
	AT	3	-			
<i>T. kodakarensis</i> KOD1	T	31	-	41	12	-
	AT	23	-			
<i>E. coli</i> str. K-12 substr. MG1655	T	11	2	12	4	-
	AT	4	-			

*asRNAs anotados nesse estudo e encontrados por Lybecker et al. (2014).

Em *H. salinarum* NRC-1, cinco dos novos genes anotados como constituintes de TAs do tipo II estavam previamente anotados como proteínas hipotéticas e os demais com outras funções (Apêndice A – Materiais suplementares - S7). Dentre esses, destacamos o par de genes VNG_RS11890 (toxina) e VNG_RS11895 (antitoxina). Identificamos asRNAs para ambos (Figura 31), A expressão da fita sense da toxina é quase imperceptível, em relação a de seu asRNA.

Figura 31- asRNAs anotados em um sistema TAs do tipo II em *H. salinarum* NRC-1. Os genes VNG_RS11890 (toxina) e VNG_RS11895 (antitoxina) (setas laranças, codificados na fita *reverse*) constituem um sistema TAs do tipo II. Os asRNAs anotados para ambos são representados pelas setas rosas. Em vermelho, no painel superior o perfil do sinal dos dados

de RNA-seq (log2 da contagem de reads por posição) para fita *forward* e inferior para a *reverse*, Triângulos pretos representam TSSs.

4.8-Expressão diferencial na curva de crescimento

Para avaliar os asRNAs diferencialmente expressos e que podem possivelmente atuar em *cis*, realizamos uma análise de expressão diferencial utilizando bibliotecas de RNA-seq geradas a partir de amostras de RNA coletadas em três pontos representativos da curva de crescimento de *H. salinarum* NRC-1 (Figura 32). Os pontos foram coletados as 17 horas (17h), representando a primeira fase de crescimento; as 37 horas (37h), início da fase de crescimento exponencial e as 86 horas (86h), quando o crescimento está estagnado (Facciotti et al., 2010). Avaliamos os asRNAs e seus genes cognatos diferencialmente expressos nas passagens de 17H para 37H e 37H para 86H.

Figura 32- Pontos de amostragem na curva de crescimento. Os pontos indicam os momentos da curva em que os *pellets* foram coletados em réplicas biológicas para extração de RNA. Correspondem a 17H, 37H e 86H (Extraído de: TEN-CATEN et al., 2018).

Na transição de 17H para 37H observamos 93 asRNAs diferencialmente expressos (Apêndice A – Materiais suplementares - S12), 75 *downregulados* e 18 *upregulados*. Para 26 desses, o gene cognato também está diferencialmente expressos (Figura 33). 15 pares apresentam correlação positiva dos níveis de expressão e 11 negativa. Três pares possuem interação com LSm em ambos transcritos, e um par com apenas o asRNA interagindo. Seis asRNAs sobrepõe regiões da extremidade 5'UTR dos genes.

Figura 33- Pares asRNAs/genes diferencialmente expressos na passagem de 17H para 37H. Valores representam o log2 do *fold change* do gene ou do asRNA em relação ao ponto anterior da curva de crescimento. Em tons de vermelho genes *upregulados* e de verde genes *downregulados*.

Na transição de 37H para 86H observamos 63 asRNAs diferencialmente expressos (Apêndice A – Materiais suplementares - S13), sete *downregulados* e 56 *upregulados*. Para 30 desses, o gene cognato também está diferencialmente expressos (Figura 34). 27 pares apresentam correlação positiva dos níveis de expressão, todos com asRNA e gene *upregulados*, e apenas três apresentam correlação negativa. Para todos os 13 asRNAs

interagindo com LSm, o transcrito do gene também está interagindo, todos *upregulados*. Sete asRNAs sobrepõe regiões da extremidade 5'UTR dos genes.

Figura 34- Pares asRNAs/genes diferencialmente expressos na passagem de 37H para 86H. Valores representam o log2 do *fold change* do gene ou do asRNA em relação ao ponto anterior da curva de crescimento. Em tons de vermelho genes *upregulados* e de verde, genes *downregulados*.

No total, encontramos 144 asRNAs diferencialmente expressos em, ao menos, uma das duas transições de fase analisadas (Apêndice A – Materiais suplementares - S14). 28 desses asRNAs interagem com LSm, 38 sobrepõem a extremidade 5' dos genes cognatos. 49

desses asRNAs, são encontrados em genes que codificam proteínas hipotéticas, e 16 em transposases. Ao todo, 56 pares gene/asRNA se encontram diferencialmente expressos em ao menos uma das duas transições analisadas (Figura 35). Os asRNAs desses pares são potenciais reguladores em *cis* da expressão de seus genes cognatos. Desses, um total de 16 asRNAs está interagindo com LSm e apenas para um desses o transcrito do gene sense não interage com essa proteína.

Figura 35- 56 pares asRNA/gene diferencialmente expressos em ao menos uma das duas transições da curva de crescimentos analisada. Todos os gráficos de dispersão apresentam os valores de $\log_2$ do *fold change* dos genes no eixo X e dos asRNAs no eixo y para os 56 pares encontrados diferencialmente

expressos. As características dos pares indicados pelas cores em cada gráfico estão explicitadas para cada um.

4.9-aTSSs conservados em *H. volcanii*

Avaliamos os possíveis aTSSs conservados entre *H. salinarum* NRC-1 e *H. volcanii*. Geramos uma lista de pares de genes ortólogos entre esses dois organismos utilizando a base de dados OrtholugeDB 2.1 (disponível em <http://www.pathogenomics.sfu.ca/ortholugedb/>) (WHITESIDE et al., 2013) e encontramos 1554 pares de genes ortólogos entre os dois halófilos. Geramos uma lista dos genes ortólogos que possuem aTSSs nos dois organismos. Ao todo 434 pares de ortólogos entre *H. salinarum* NRC-1 e *H. volcanii* possuem aTSSs. Quando buscamos apenas pelos genes em que anotamos um asRNA em NRC-1, obtivemos 244 pares de ortólogos, ou seja, que possuem um asRNA anotado em *H. salinarum* NRC-1, e ao menos um aTSS em *H. volcanii*. Dentre esses últimos, buscamos por aTSSs em *H. volcanii* que se encontram na mesma região do gene em que se encontra um aTSSs que dá origem a um asRNA em *H. salinarum* NRC-1. Os genes tiveram seus tamanhos normalizados e divididos em 10 regiões.

Ao todo, 160 dos asRNAs inferidos em *H. salinarum* NRC-1, distribuídos em 110 genes, possuem ao menos um aTSS correspondente na mesma região de seu gene ortólogo em *H. volcanii* (Apêndice A – Materiais suplementares - S15). Para esses 110 genes, filtramos aqueles que possuem exatamente o mesmo tamanho em nt nos dois organismos para avaliarmos quantos aTSSs possuem uma conservação da posição exata em que são encontrados (Apêndice A – Materiais suplementares - S16). Os pares ortólogos VNG_RS10385 e HVO_0359, que codificam uma subunidade de um fator de alongamento da tradução e o par VNG_RS00180 e HVO_1531, que codificam uma UDP-glucose desidrogenase, possuem aTSSs com conservação da posição exata nos genes ortólogos dos dois organismos (Figura 36). Todos apresentam uma guanina (G) na posição do aTSS. Analisamos a porcentagem de nucleotídeos conservados na região referente ao asRNA anotado em *H. salinarum* NRC-1, obtendo 89,4% para o asRNA do gene VNG_RS10385 e 67,3% para o asRNA do gene VNG_RS00180.

Figura 36- Exemplos de aTSSs conservados em *H. salinarum* NRC-1 e *H. volcanii*. Dois pares de genes ortólogos que possuem mesmo tamanho em nt e aTSSs exatamente na mesma posição. A direita da representação dos genes dos pares ortólogos estão os alinhamentos múltiplos das sequências correspondentes a região do asRNA anotado em *H. salinarum* NRC-1.

4.10-RNase R

Comparamos as sequências de aminoácidos (aa) das RNases R codificadas por *H. salinarum* NRC-1 (gene VNG2647G) e *H. volcanii* (gene HVO_0388) para verificar a existência de homologia. Tal como em *H. volcanii*, o gene que codifica essa RNase em *H. salinarum* NRC-1 possui cópia única no genoma. A sequência de aa de ambas possui tamanho similar, 429aa para a de *H. salinarum* NRC-1 e 439aa a de *H. volcanii*. Utilizando a ferramenta BlastCD, identificamos os domínios proteicos presentes nas RNases R dos dois organismos. As duas apresentam um único e mesmo domínio, “VacB” (Figura 37). O alinhamento múltiplo dessas sequências mostra que a porcentagem de identidade total entre elas é de 68,58%, e quando consideramos apenas a região do domínio, 74.92% (Figura 38). Tais resultados nos permite inferir que se tratam de proteínas homólogas.

Figura 37- Resultado da busca por domínios proteicos nas RNases R utilizando BlastCD. Ambas sequências de aa apresentam um único domínio proteico, “VacB”. A barra rosa representa a região relativa ao domínio na sequência de aa. Primeiro resultado referente a sequência de *H. salinarum* NRC-1 (429 aa) e segundo a de *H. volcanii* (439aa).

1 10 20 30 40 50

H. salinarum .MSDSEPAQGSIEDQGPVEISP ELARQL ENKRDDLFEKFDLHDEFPPDVVAEAE
H. volcanii MSDDAQAYA GTAE GQGPVEIDAE ELARHLQ NKREEELFE EFEIRD AFPEVLSEAR

60 70 80 90 100

H. salinarum ARTEDEVQA EIQAEIEDRADMRPLPTWTVDPVDAQDFDDAISVLERDDEYVLWVH
H. volcanii ARTDGVHEIEIED ELEHRKDLRLDTWTD PVDAQDFDDALSIRENEEYTLWVH

110 120 130 140 150 160

H. salinarum IADVTHYVHPD SEMWAEAIID RANTVYLPGYT M HMLPPVLAETVCSLPVGEDRLA
H. volcanii IADVTHYVHPG SEMWAEAVKRG NTVYLPAYT I HMLPPALAEETVCSLPVEEDRLA

170 180 190 200 210

H. salinarum HTVEMHLDKDDLSYQTIEIYKSVIHS DERLTYKDAEQRLDDETAAPLHDDLSLVLF
H. volcanii HTVEMEIKKDTLSFESID IYKSVIHSNERLTYTQCENRIEDPDLPLHEENALVLF

220 230 240 250 260

H. salinarum DLADQMHEQRKA EGSLVLNPR RDR AHTIIEES MLKANKAVTHELMWNRGVEAVY
H. volcanii DLADQLHEQRKE DGS LVLNPS RDR AHTIIEEC MLKANKAVTHELMWNRGVEAMY

270 280 290 300 310 320

H. salinarum RVHPQPSPEEWD DALQEIQELN GVSIPGD AWD D PRKAVNATLEQAPERQLGKIQ
H. volcanii RVHPQPTPDQWD KALREITELD GVSIQSTSWDEPRKAVNDALESANSTR LNKIQ

330 340 350 360 370

H. salinarum WAVMKVMPRAR YMNDFEGGH H ALNFEIYGHFTSPIRRMSDI VNHWIVYQNDVPE
H. volcanii RAVLKVMPRAK YMNDFEGGH Y ALNFDIYGHFTSPIRR LSDI INHWIVHENNDVPE

380 390 400 410 420

H. salinarum T L I S L CDRASD K Q A G E T A E R E Y R D F L E E V G L D P D A V N N R C V T V V D S P T T S S . .
H. volcanii D L V E L CDRASD R Q K D A E T A E R L Y K Q F L Q E V G L D P Y A V N N R C I V T V D D D G E V I D E

H. salinarum
H. volcanii NGLPPOE

Figura 38- Alinhamento múltiplo das sequências de aa das RNases R de *H. salinarum* NRC-1 e *H. volcanii*. Resíduos idênticos em vermelho e similares em amarelo.

Obtivemos uma linhagem *knockout* para o gene VNG2647G, fornecida pelo Laboratório do Professor Nitin Baliga (*Institute for Systems Biology*). A deleção do gene foi confirmada por PCR. Verificamos o fenótipo do mutante por ensaio de curva de crescimento padrão (Figura 39). Além de viável, a partir das 50 horas da curva de crescimento, a linhagem *knockout* apresenta maior crescimento do que a linhagem controle.

Figura 39- Curva de crescimento. Avaliação do crescimento da linhagem *knockout* do gene VNG2647G em comparação com a linhagem controle. Pontos representam médias das leituras de três réplicas biológicas. Barras representam o desvio padrão das leituras das réplicas por ponto.

Com a possibilidade de realizar estudos com a linhagem *knockout* viável, realizamos um RNA-seq da linhagem *knockout* e comparamos com a linhagem controle. Baseados no relato de atividade de degradação de dsRNAs de (MATOS, RUTE G. et al., 2012), nossa hipótese é de que para os genes que possuem asRNAs e que possivelmente seus mRNAs formam duplex de RNA com essas moléculas, conseguiríamos observar um aumento simultâneo na expressão do asRNA e do mRNA na linhagem *knockout*. A lista completa dos genes diferencialmente expressa se encontra em Apêndice A – Materiais suplementares – S17 e S18. Apêndice

Encontramos apenas cinco asRNAs diferencialmente expressos na linhagem *knockout*, três *downregulados* e dois *upregulados* (Tabela 8), sendo que apenas para um desses, o gene sense também foi encontrado diferencialmente expresso. Esse único par asRNA/gene diferencialmente expresso se encontra *upregulado*, um perfil que esperávamos encontrar para vários casos se a RNase R tivesse um papel essencial na degradação de dsRNAs *in vivo* tal como a RNase III em bactérias (LASA, I et al., 2011) (LYBECKER et al., 2014). O par em questão é constituído pelo gene VNG_RS03830 ($\log_2FC = 2.75$) que codifica o tRNA Lys e um asRNA de 133nt ($\log_2FC = 3.06$) que sobrepõe todo o gene sense (Figura 40),

Tabela 8: asRNAs diferencialmente expressos na linhagem *knockout* para o gene VNG2647G.

asRNA	log2FC	Gene sense	anotação gene sense
aTSS_20	-2.35	VNG_RS00105	transposase
daTSS_82	-2.12	VNG_RS13545	hypothetical protein
aTSS_2056	-1.66	VNG_RS11690	IS4 family transposase ISH30
daTSS_44	2.98	VNG_RS06840	Nif3-like dinuclear metal center hexameric protein
daTSS_26	3.06	VNG_RS03830	tRNA-Lys

Figura 40- Par asRNA/gene *upregulado* na linhagem *knockout* para o gene VNG2647G. O gene VNG_RS03830 (seta amarela, codificado na fita *forward*) codifica o tRNA Lys. O asRNA (seta rosa) anotado para esse gene inicia a partir de um daTSS e sobrepõe todo o gene. Em verde, perfil do sinal dos dados de RNA-seq ($\log_2$ da contagem de *reads* por posição) das duplicatas da linhagem *knockout* e em rosa da linhagem controle. Triângulo preto indica o daTSS.

5-Discussão

5.1-Mapeamento de aTSSs e anotação de asRNAs: mapeamento das extremidades 5' e inferência das extremidades 3'

A técnica dRNA-seq tem permitido a identificação de transcritos primários sobrepostos a regiões codificadores, principalmente antisense, em bactérias e arqueias (SHARMA; VOGEL, 2014)(GELSINGER; DIRUGGIERO, 2018a). Nesse estudo, mapeamos 1626 aTSSs no genoma de *H. salinarum* NRC-1. Para outras arqueias que possuem dados de dRNA-seq publicados, um número menor, porém não distante de aTSSs foi mapeado, sendo, 1244 para *H. volcanii* (BABSKI et al., 2016), 1110 para *Methanobrevibacter smithii* (LI, JIE et al., 2015) e 1018 para *Thermococcus kodakarensis* (JÄGER et al., 2014). Em *H. volcanii*, dessas a arqueia filogeneticamente mais próximo de *H. salinarum* NRC-1, a porcentagem de genes em que foi possível encontrar aTSSs durante crescimento exponencial e condições ótimas de crescimento é de ~30%, similar a porcentagem de genes com aTSSs que obtivemos, ~34%. Para *H. volcanii*, a análise da distribuição da posição dos aTSSs em relação ao gene sense obteve resultado semelhante ao nosso mostrado na Figura 17A, onde um acúmulo de aTSSs na extremidade 3' dos genes pode ser observado. Os 1626 aTSSs mapeados nesse estudo podem dar origem a moléculas de RNA que atuarão pós-transcricionalmente no controle da expressão de seus genes cognatos por pareamento de bases com os mRNAs produzidos por tais genes. Para verificar essa hipótese, integramos diversas bibliotecas de RNA-seq de *H. salinarum* NRC-1 para buscar transcritos que começam exatamente onde um aTSS foi mapeado.

O aumento da resolução na identificação de novos transcritos foi um dos grandes avanços permitidos pela técnica RNA-seq. Essa técnica é sensível a pequenas variações nos níveis de expressão, permitindo identificar variações dos níveis de transcrição de regiões adjacentes e sobrepostas (WANG, 2009). Inferir com precisão o tamanho de um transcrito, ou seja, suas extremidades 5' e 3', ainda é um desafio, pois mesmo com o grande aumento de resolução proporcionado em relação as técnicas de avaliação de transcrito que a antecederam, o transcrito ainda tem de ser montado a partir de sequências pequenas (CONESA et al., 2016; GARBER et al., 2011). Em procariotos, baseada no fato de que transcritos primários possuem suas extremidades 5' trifosfatadas, a técnica dRNA-seq permite a identificação global e automatizada das extremidades 5' de transcritos primários

com alta resolução (SHARMA; VOGEL, 2014). Porém, o término desses transcritos (extremidade 3') não é identificado de forma automática. Nos estudos utilizando dRNA-seq para analisar os transcriptomas primários de *E. coli* (THOMASON et al., 2015) e *H. volcanii* (BABSKI et al., 2016), os autores delimitaram arbitrariamente o tamanho dos transcritos associados aos aTSS mapeados em 50 e 100 nt respectivamente, uma abordagem simplista para realizar generalizações sobre a expressão dos asRNAs associados aos aTSS mapeados. Uma outra abordagem possível é a utilização de assinaturas intrínsecas ao término de transcrição para inferir o final dos transcritos de forma automatizada. Para a cianobactéria *Nostoc sp.* PCC 7120, (BRENES-ÁLVAREZ et al., 2016) dados de TSS obtidos por dRNA-seq foram combinados com assinaturas de término de transcrição Rho-independentes preditas para o genoma desse organismo para inferir a extremidade 3' de ncRNAs. Essas assinaturas indicam, por exemplo, a formação de estruturas no final do transcrito que encerrarão a transcrição, e portanto, não dependem da proteína Rho para isso (RAY-SONI; BELLECOURT; LANDICK, 2016). Tal critério baseado em informações a priori sobre a terminação de transcrição, oferece um melhor embasamento biológico em relação ao simples estabelecimento arbitrário do tamanho dos transcritos.

Para maioria dos transcritos em bactérias, a terminação da transcrição ocorre de forma independente de fatores de terminação. Geralmente, as sequências do final dos transcritos são ricas em uracila. A transcrição dessas regiões leva a pausa da RNA polimerase e promove a formação de *hairpins*, levando a desestabilização do complexo da RNA polimerase e, conseqüentemente, ao término da transcrição (GUSAROV; NUDLER, 1999; PETERS; VANGELOFF; LANDICK, 2011). Porém, nem todos organismos possuem seus mecanismos de término de transcrição elucidados, impedindo a mesma abordagem utilizada por Brenes-Álvarez et al. (2016) na inferência do término dos transcritos. Para alguns organismos do domínio Archaea, o término da transcrição pode estar associado a sequências ricas em uracila, porém, se comparado a bactérias, os mecanismos de término de transcrição ainda são pouco compreendidos nesse grupo, dificultando o estabelecimento de assinaturas de sequências precisas para determinar o fim dos transcritos (DAR et al., 2016; SANTANGELO; REEVE, 2006).

A anotação de transcritos a partir da inspeção visual do decaimento de sinal da expressão gênica em dados de RNA-seq para inferir as extremidades 3' é uma abordagem já empregada para a anotação de ncRNAs, incluindo asRNAs, em bactérias e arqueias (GELSINGER; DIRUGGIERO, 2018b; IRNOV et al., 2010; VAN DER MEULEN; DE JONG; KOK, 2016; WURTZEL et al., 2010b). Para anotar os finais dos transcritos associados aos aTSS mapeados em *H. salinarum* NRC-1, utilizamos um recurso técnico aliado a curagem manual integrando diferentes dados. Utilizamos dados de bibliotecas de RNA-seq fita específicas sequenciadas em modo *paired-end* para curagem manual dos transcritos associados aos aTSS mapeados. O sequenciamento em modo PE gera sequências das duas extremidades de um mesmo transcrito, dessa forma, mesmo sem a possibilidade de sequenciar um transcrito cujo tamanho seja maior que os reads, suas extremidades podem ser inferidas após alinhar os pares de reads. Essa técnica é amplamente utilizada na montagem de genomas e de transcriptomas de eucariotos. Por sequenciar os transcritos em duas direções, pressupomos que esses dados nos dariam maior confiabilidade para estabelecer o final dos transcritos, de modo que o final daqueles maiores que o tamanho dos reads também podem ser identificados e mesmo nos casos em que o tamanho dos transcritos é menor que os reads, o sequenciamento em ambas direções pode garantir maior confiabilidade para a inferência de suas extremidades. A Figura 19 mostra um exemplo de como a utilização de dados PE permite uma anotação mais acurada das extremidades 3' de um transcrito.

Verificamos se nosso método de anotação de asRNAs poderia nos mostrar algum padrão de sequência na extremidade 3' dos transcritos que pudesse indicar uma sequência terminadora (Figura 22). Comparado as assinaturas poli-U de terminação já encontradas em arqueias, esse resultado mostra uma região muito curta, sendo pouco provável que tenha algum tipo de influência na formação de uma estrutura terminadora (DAR et al., 2016). Porém, a observação de um enriquecimento de uracilas exatamente na última e penúltima posição base dos transcritos é algo que chama a atenção, podendo indicar alguma assinatura de terminação.

A extensão de um transcrito com a sua extremidade 3' podem ser obtidos com precisão utilizando técnicas como NorthernBlot (quando a extremidade 5' é conhecida) e 3'RACE, porém se tratam de técnicas laboriosas que normalmente exigem muito tempo de

padronização, sendo inviáveis para estudos como esse, cuja intenção é avaliar o panorama global de uma classe de ncRNAs em um organismo. Em muitos estudos desse tipo, tais técnicas são utilizadas para a validação de alguns ncRNAs em específico (LYBECKER et al., 2014; THOMASON et al., 2015; VAN DER MEULEN; DE JONG; KOK, 2016). A detecção experimental em larga escala de extremidades 3' de transcritos ainda consiste em um desafio metodológico se comparados com a relativa simplicidade do dRNA-seq. Variações no preparo de bibliotecas de RNA associadas ao sequenciamento em larga escala das extremidades das moléculas de cDNA geradas também tem sido utilizadas para esse propósito. São exemplos: RACE-seq (rapid amplification of cDNA ends) (SALEHI-ASHTIANI et al., 2009; GOLDFARB; CECH, 2013), *high-throughput circularization protocols* (PELECHANO et al., 2013), TIF-Seq e Term-seq (DAR et al., 2016). Outro método em larga escala que recentemente se tornou viável é o sequenciamento de transcritos inteiros sem necessidade de fragmentação ou produção de cDNA no preparo da biblioteca utilizando o sequenciador MiniION (GARALDE et al., 2018; JENJAROENPUN et al., 2018).

Em nosso laboratório, a integração de dados em larga escala como RNA-seq e tilingarray associada a curagem manual de sinais de transcrição é uma abordagem inicial que tem permitido a descoberta de novas classes de RNAs conservados nos três domínios da vida. Além disso, para os casos em que evidências experimentais foram obtidas, as anotações dos transcritos obtidas a partir de dados em larga escala se mostraram concordantes, evidenciando um grande poder de resolução das técnicas utilizadas para a obtenção dos dados (GOMES-FILHO et al., 2015; ZARAMELA et al., 2014). Tal histórico nos faz confiantes de que a abordagem de anotação de asRNAs, baseada na curagem de sinais de diferente bibliotecas de RNA-seq, utilizada nesse estudo é precisa o bastante para sustentar as posteriores análises de expressão gênica realizadas e, provavelmente, em muitos casos, reflete com precisão o tamanho real dos asRNAs anotados. O objetivo desse estudo foi obter um panorama do transcrito antisense primário de *H. salinarum* NRC-1, a validação *in vitro* de alguns transcritos aqui anotados e discutidos utilizando algumas das técnicas citadas poderá vir a ser um novo passo para a consolidação de hipóteses e investigação de mecanismos de regulação gênica inferidos a partir dos resultados aqui obtidos. Além da validação *in vitro*, a recuperação nos dados de larga escala de transcritos já validados em outros estudos e a presença de

transcritos conservados entre espécies, são evidências que conferem uma alta confiabilidade aos resultados, mesmo se tratando de um mapeamento em escala genômica.

Algo a ser considerado é que o esforço de curagem manual e os recursos técnicos empregados nesse trabalho para a anotação dos asRNAs oriundos dos aTSSs se restringe ao estudo do fenômeno de regulação antisense pós-transcricional, pois, quando buscamos por sinais de transcritos constantes em várias bibliotecas e a definição de uma posição 3', assumimos o pressuposto de que esses aTSSs compreendem inícios de *loci* de ncRNAs funcionais e expressos sempre com tamanho constatante. Porém, nada impede que os demais aTSSs mapeados, para os quais não foram anotados asRNAs, também participem de alguma forma em outros estágios da regulação pós-transcricional. O mapeamento dos aTSSs não esgota todas as possibilidades de regulação antisense que pode ocorrer em um organismo. Uma análise prévia realizada em nosso laboratório mostrou que 85% dos genes anotados para *H. salinarum* NRC-1 possuem sinal de transcrição antisense cobrindo pelo menos metade de suas extensões. No presente estudo encontramos ao menos um aTSS para 43% dos genes anotados e ao menos um asRNAs para 21%. Um exemplo de fenômeno de geração de sinal antisense não relacionado ao transcrito antisense primário que pode estar relacionado a grande quantidade de genes com sinal antisense são os mRNAs com 3'UTR longos (HERNÁNDEZ et al., 2006; RASMUSSEN; NIELSEN; JARMER, 2009). Na Figura 41, apresentamos um exemplo desse caso que encontramos ao longo da anotação dos asRNAs. Além disso, a regulação pós-transcricional também pode ocorrer mediada por asRNAs não primários, cujo início de transcrição não seria mapeado por dRNA-seq.

Figura 41- 3'UTR antisense longo. Seta rosa indica uma possível região 3'UTR do mRNA do gene VNG_RS04945 (seta amarela, fita *forward*, proteína hipotética) que pode atuar como asRNA para o gene VNG_RS04950 (seta laranja, fita *reverse*) que codifica a subunidade beta de um fator de iniciação de tradução. Triângulos pretos indicam TSSs. Painel superior mostra os dados de expressão gênica referentes a fita *forward*, painel inferior dados referentes a fita *reverse*. Em vermelho o perfil do sinal dos dados de RNA-seq ($\log_2$ da contagem de reads por posição). Os heat maps mostram dados de tiling array na curva de crescimento (KOIDE et al, 2009), coloridos de acordo com $\log_{10}$ da razão entre o ponto da curva em relação ao tempo zero.

5.2- asRNAs sobrepostos a extremidade 5' dos genes

A maioria dos sRNAs funcionalmente caracterizados em procariotos se liga a porção 5' do mRNA, geralmente cobrindo o códon de início da tradução (AUG) e a região RBS (*ribosome binding site*) localizado no 5'UTR (WAGNER, E GERHART H; ROMBY, 2015). Tal pareamento pode inibir o acesso da maquinaria ribossomal ao mRNA inibindo sua tradução. AsRNAs que sobrepõem essas regiões são prováveis candidatos a atuarem em *cis* por esse mecanismo por apresentarem pareamento de bases completo com essas regiões. Um

exemplo de asRNA que atua dessa forma é o já mencionado *symR*, integrante de um sistema toxina-antitoxina do tipo I em *E. coli* (Figura 11B) (KAWANO; ARAVIND; STORZ, 2007). Além disso, os dsRNAs identificados no experimento de Lybecker et al. (2014) mostram que na maioria dos pares mRNA/asRNA encontrados o asRNA sobrepõe ou está próximo a extremidade 5' do mRNA. O primeiro sRNA com seus alvos caracterizados em arqueias, sRNA₁₆₂ (PERNITZSCH et al., 2012; PRASSE; EHLERS; SCHMITZ, 2013), também atua se sobrepondo a extremidade 5' de seus mRNAs alvos, tal como a maioria dos sRNAs conhecidos em bactérias. Arqueias metanogênicas sabidamente possuem transcritos com longos 5'UTRs, de modo que essas regiões podem ser alvos de sRNAs, algo que permite inferir que sRNAs encontrados nesses organismos poderiam atuar de forma similar aos bacterianos (BABSKI et al., 2014; JÄGER et al., 2009; LI, JIE et al., 2015). Porém tal inferência não pode se estender a todo domínio Archaea, pois muitas arqueias possuem a maioria de seus transcritos *leaderless*, sendo o caso dos halófilos *H. salinarum* NRC-1 e *H. volcanii* (BABSKI et al., 2016; BRENNEIS et al., 2007). Tal característica indica que a regulação por sRNAs nesses organismos com transcritos *leaderless* ocorreria predominantemente por pareamentos na região interna das ORFs ou nas regiões 3'UTR (BABSKI et al., 2014), o que explicaria o perfil de acúmulo de aTSSs próximos a região 3' dos genes que encontramos em *H. salinarum* NRC-1 (Figura 17B) e se repete para *H. volcanii*. Exemplos de sRNAs conservados e interagindo com a região 3' dos transcritos, pareando com a região 3' UTR e/ou sobrepondo o códon de parada são reportados para arqueias conhecidas por possuírem transcritos *leaderless* (BABSKI et al., 2014). Um desses exemplos compreende arqueias do gênero *Pyrobaculum*, nos quais um asRNAs, identificado como *asR3*, de 63 nt e conservado em quatro organismos desse gênero, atua em *cis* em relação gene *tpi* (que codifica triose-phosphate-isomerase), sobrepondo o códon de parada e uma porção do 3'UTR do mRNA (LOWE et al., 2012).

Apesar de todas essas considerações feitas da relação entre a posição dos aTSSs e mRNAs *leaderless*, quando analisamos as posições relativas dos aTSSs para os quais anotamos um locus de asRNA, ou seja, os que julgamos que provavelmente estão produzindo um asRNA que atuará pós-transcricionalmente, o acúmulo de aTSSs é deslocado para a região mais próxima da extremidade 5' dos genes (Figura 17).

Bouvier et al. (2008) mostraram experimentalmente que ncRNAs que sobrepõem nucleotídeos correspondentes até o quinto códon localizado no mRNA são capazes de impedir a ligação da maquinaria ribossomal e inibir a tradução. Dos 846 *loci* de asRNAs definidos nesse estudo, 145 asRNAs sobrepõem a extremidade 5' ou no mínimo o 12º nucleotídeo do gene sense (Figura 23). Esse conjunto de asRNAs representa aqueles que, tal como sRNAs conhecidos em bactérias, podem atuar impedindo a tradução do mRNA ocluindo a ligação da maquinaria ribossomal.

5.3-Expressão em relação ao gene sense

O baixo nível de expressão é uma característica inerente a maioria dos transcritos antisense identificados em eucariotos e procariotos, sendo que em alguns organismos estudados, apresenta em média uma expressão dez vezes menor que a de seu gene cognato (LLORÉNS-RICO et al., 2016; PELECHANO; STEINMETZ, 2013; THOMASON et al., 2015). Provavelmente, esse é um dos fatores que impõe grande dificuldade na detecção desses transcritos até que métodos em larga escala de alta sensibilidade como o RNA-seq se tornassem viáveis (PELECHANO; STEINMETZ, 2013). Em nosso estudo, encontramos apenas 113 asRNAs com expressão maior ou igual à do gene cognato. Tais níveis de expressão podem ser um reflexo dos níveis reais desses transcritos na célula, sendo a maioria desses expressos em baixa quantidade ou alvos de rápidos processos de degradação por RNases; ou de uma grande instabilidade que torna esses transcritos suscetíveis a degradação durante o isolamento de RNA e preparo de bibliotecas para o sequenciamento (THOMASON et al., 2015). Se a formação de um dsRNA é necessária para que ocorra algum fenômeno de regulação pós-transcricional, tomando considerações estequiométricas básicas, é esperado que para que haja efeito regulador, um asRNA seja encontrado em proporções equivalentes à de seu RNA alvo na célula. Tal consideração leva a hipótese de que asRNAs funcionais teriam níveis de expressão equivalentes aos de seus genes alvos, sendo esse o caso para alguns asRNAs com funções caracterizadas em bactérias (FOZO; HEMM; STORZ, 2008). Lloréns-Rico et al. (2016) apresentam evidências de que muitos asRNAs bacterianos podem ser meros transcritos espúrios, oriundos de regiões promotoras de baixa complexidade com alta possibilidade de se originarem aleatoriamente em um genoma. Os genes desses asRNAs espúrios seriam mantidos nos genomas pela ausência de seleção negativa contra esses loci

gênicos, provavelmente por conta do baixo custo energético na produção desses transcritos e ausência de efeitos deletérios. Porém, os autores desse estudo ressaltam que a detecção da baixa expressão de um asRNA não exclui a possibilidade de possuírem função biológica. Pouco se sabe da interação desses RNAs com proteínas que poderiam mantê-los funcionais por longos períodos de tempo no citoplasma, tendo grande impacto na regulação mesmo em baixas quantidades. Além disso, mesmo em baixas quantidades, asRNAs podem ter um efeito “tamponante”, atuando como reguladores finos da expressão de seus mRNAs, de modo que, impediriam a transcrição espúria desses últimos e impedindo o gasto energético desnecessário da produção de proteínas, criando um limiar, do qual a tradução só ocorreria quando o mRNA superasse o nível de asRNAs (Figura 11D) (LASA et al., 2011; PELECHANO; STEINMETZ, 2013).

Ainda que expressos em baixas quantidades e talvez, não funcionais, muitos dos asRNAs identificados por métodos em larga escala, inclusive os nesse estudo em *H. salinarum* NRC-1, podem ser detectados em diferentes condições biológicas e são expressos constantemente com o mesmo tamanho e perfil de sinal composto por reads. Esses transcritos podem servir de matéria-prima para processos evolutivos que podem originar novos elementos reguladores funcionais para o organismo. Tais transcritos podem vir a se tornar “exaptações” moleculares. Exaptação é um conceito associada a uma perspectiva neutralista da evolução molecular, foi cunhado por Brosius e Gould (1992) para se referir a características genômicas que surgem aleatoriamente ou por pressões seletivas não relacionadas ao papel funcional que desempenham, o que poderia se aplicar ao surgimento de novos asRNAs reguladores a partir de eventos espúrios transcrição.

5.4-Região Promotora

Na Figura 25, pode ser notada a predominância de purinas na posição 0, correspondente a posição do TSS e de pirimidinas na posição -1. Tal resultado é concordante com o que é conhecido sobre início de transcrição em arqueias halofílicas e outras arqueias como *Sulfolobus solfataricus* (WURTZEL et al., 2010b) e *Pyrobaculum aerophilum* (SLUPSKA et al., 2001), além de ser uma característica também conservada em eucariotos (CARNINCI et al., 2006; ZHANG, ZHIHONG; DIETRICH, 2005) e bactérias (KIM, DONGHYUK et al., 2012). Há evidências de que essa composição dinucleotídica é

importante para o reconhecimento da posição de início da transcrição, de modo que, alterações nesses sítios podem diminuir a expressão do gene. Também é possível observar assinaturas de sequências correspondentes a elementos característicos de regiões promotoras em arqueias, sendo esses uma TATA-box na posição -26, e o sítio de reconhecimento do fator de transcrição TIIB (B recognition element – BRE) caracterizado pelo enriquecimento de nucleotídeos GC na posição -36 (BELL; JACKSON, 2001; SEITZER et al., 2012). Tal resultado evidencia que o dRNA-seq, como método de mapeamento em larga escala de TSSs, gera resultados coerentes com dados prévios obtidos por outros métodos experimentais e análises, evidenciando o poder de resolução e confiabilidade dessa técnica para a descoberta de novos transcritos.

5.5- asRNAs que interagem com LSm

Análises prévias do nosso grupo mostram que um total de 415 genes anotados em *H. salinarum* NRC-1 apresentam alguma região de interação com LSm (~15% dos genes anotados). No presente estudo, identificamos um total de 91 asRNAs com região de interação com LSm. 48 desses asRNAs pertencem a genes que codificam proteínas hipotéticas. 7 dos asRNAs interagindo com LSm pertencem a genes que codificam elementos transponíveis.

Em *H. volcanii*, foram detectados 20 sRNAs co-purificados com a proteína LSm (FISCHER et al., 2010). Desses, dois RNAs antisense a elementos transponíveis, RNA 25 antisense ao gene HVO_C0039 e o RNA 144, antisense ao gene HVO_0263, ambos genes classificados como ISH51 (STRAUB et al., 2009). Para o transposon composto bacteriano Tn10/IS10, Ross et al. (2013) obtiveram evidências a partir de experimentos *in vitro* de que o asRNA de 69nt, pertencente ao gene IS10R, interage com a proteína Hfq em um mecanismo de repressão *in cis* bloqueando o início de tradução localizado na extremidade 5' do mRNA desse gene (ROSS et al., 2013). asRNAs advindo de elementos transponíveis tem sido reportados em vários estudos para diferentes arqueias (GELSINGER; DIRUGGIERO, 2018a), evidenciando uma possível relação funcional entre asRNAs e eventos de transposição nesses organismos. Resultados recentes de nosso grupo indicam que em *H. salinarum* NRC-1, aproximadamente 63% das IS intactas produzem RNAs que apresentam interação com LSm, sendo proporcionalmente mais propensos a interação com ela do que os RNAs derivados dos demais genes. Tais resultados somados aos asRNAs advindos desses elementos transponíveis

interagindo com essa proteína aqui reportados, compõem um importante ponto de partida para a compreensão da relação entre LSm e asRNAs no controle de eventos de mobilização em genomas de arqueias.

Dos 846 asRNAs anotados, uma pequena fração, 81 (~9,6%), podem constituir complexos asRNA-Lsm-mRNA, algo que quando relacionado a proteína Hfq, é predito como um processo que reduzirá a energia de ativação para a formação do duplex que de alguma forma, pode estar relacionado a um fenômeno de regulação pós-transcricional (VOGEL, JÖRG; LUISI, 2011).

Algo a ser considerado, é que, muito do que se conhece da interação de ncRNAs com seus alvos mediada por Hfq em bactérias e complexos proteicos em eucariotos está relacionado a ncRNAs que atuam em *trans*, tais RNAs tem por característica o não pareamento completo com seu alvo e serem codificados em regiões genômicas distintas de seus alvos (WILUSZ; WILUSZ, 2005). Tais características reduzem a probabilidade de encontro dessas moléculas com seus alvos no citoplasma e posterior ligação por complementariedade de bases, justificando a necessidade de complexos proteicos ou proteínas mediando a formação de um duplex ncRNA/RNA-alvo. Para RNAs antisense que atuam em *cis*, esses dois obstáculos que, quando quantificados em termos de termodinâmica de interação de moléculas, contribuirão para a energia de ativação que leva a formação do complexo, constituem barreiras transpostas com maior facilidade termodinâmica (VOGEL, JÖRG; LUISI, 2011). O pareamento sempre será completo e as moléculas são sintetizadas fisicamente próximas a seus alvos, levando a inferir que em relação aos ncRNAs que atuam em *trans*, a interação com uma proteína seria de menor importância para os asRNAs (HOLMQVIST; WAGNER, 2017; WATERS; STORZ, 2009). Para a bactéria patogênica *Salmonella enterica*, há evidências de que muitos sRNAs que sabidamente não interagem com a proteína Hfq, interagem com a proteína ProQ, uma proteína que também classificada com uma RBP. Dentre os sRNAs encontrados interagindo com essa proteína, estão vários asRNAs (HOLMQVIST; WAGNER, 2017). Até o momento, os esforços relacionados a identificação de proteínas que interagem com sRNAs em arqueias se resumem ao estudo com *H. volcanii* (FISCHER et al., 2010) e as pesquisas em andamento do nosso grupo com a mesma proteína em *H. salinarum* NRC-1, ambos tendo como alvo a proteína LSm por se tratar de uma homóloga de Hfq. Porém, tal como observado em *S. enterica*, proteínas RBPs ainda não

caracterizadas nesses organismos podem ter papel de igual ou maior relevância que LSm no entendimento da regulação pós-transcricional por sRNAs. No terceiro domínio da vida, um campo de estudo com muito a ser descoberto.

Na Figura 26, mostramos dois exemplos de genes com asRNAs anotados e regiões de interação com LSm em transcritos correspondentes as fitas *forward* e *reverse*. Esses genes codificam glicosiltransferases da família 1 e provavelmente estão envolvidas na formação da *S-layer*, uma camada de glicoproteínas que envolve a membrana plasmática de *H. salinarum* NRC-1 (DYALL-SMITH, 2008). Ambos genes chamam a atenção por apresentarem um perfil semelhante de distribuição de asRNAs. Essa observação é uma evidência de asRNAs interagindo com LSm envolvidos no processo de formação de um importante componente estrutural celular desse organismo.

5.6-Função dos genes sense potencialmente regulados por asRNAs

Como esperado para uma classe de ncRNAs associados a um fenômeno pervasivo como a transcrição antisense, os genes nos quais os asRNAs foram identificados estão associados a uma grande quantidade de diferentes funções, muitas delas essenciais ao organismo, algo que evidencia a relevância dessa classe de ncRNAs para *H. salinarum* NRC-1 como possíveis reguladores de diferentes processos celulares.

5.6.1-Tradução

Identificamos asRNAs em vários genes relacionados ao processo de tradução. asRNAs advindos de tRNAs também foram identificados para os genes correspondentes aos Ser, The e Pro em *S. solfataricus* (TANG et al., 2005). Para *T. kodakarensis*, os autores do estudo utilizando dRNA-seq também reportam o mapeamento de aTSS em genes de tRNAs, porém não fornecem maiores detalhes sobre quais (JÄGER et al., 2014).

Esses genes relacionados ao processo de tradução e seus homólogos em organismos de todos domínios são essenciais para o funcionamento celular. A mineração de mais dados de transcriptoma de outros organismos e mais análises são necessárias para descobrir se esses asRNAs são conservados, algo que do ponto de vista evolutivo, é de grande relevância investigar se essas moléculas estão associadas a regulação de tais genes.

5.6.2-Transposases

AsRNAs presentes em transposases também foram reportados em *H. volcanii*, *T. kodakarensis*, *S. solfataricus* e *M. mazei* (GELSINGER; DIRUGGIERO, 2018a). Em *H. volcanii*, 134 dos 1244 aTSSs mapeados pertencem a genes anotados como transposase (BABSKI et al., 2016). Para *H. salinarum* NRC-1, obtivemos um resultado de similar magnitude. 155 dos 1626 aTSSs mapeados pertencem a genes anotados como transposases

Em *T. kodakarensis*, ao menos um aTSSs foi mapeado para todos os genes que codificam transposases nesse organismo (JÄGER et al., 2014). 8 asRNAs presentes em transposases foram identificados em *S. solfataricus* (TANG et al., 2005). Um dos poucos asRNAs caracterizados funcionalmente em arqueias é advindo de uma transposase de *S. solfataricus*. O RNA-257, como foi identificado, atua em *trans* pareando a extremidade 3' de um mRNA de um transportador de fosfato regulando sua expressão (MÄRTENS et al., 2013). Em *M. mazei*, 6 asRNAs foram detectados em regiões 5'UTR de transposases. Alguns desses asRNAs são diferencialmente expressos em resposta a variação de concentração de nitrogênio no meio, o que os autores do trabalho relacionam a um possível controle de transposição em meio a uma situação de estresse metabólico (JÄGER et al., 2009). Em bactérias, o asRNA art200 de 90nt, presente em uma transposase da família IS200/IS605, atua em *cis* na repressão da tradução do mRNA desse gene se ligando a região 5'UTR inibindo a ligação da maquinaria ribossomal ao mRNA, tal interação ocorre independente de interação com a proteína Hfq (ELLIS; TRUSSLER; HANIFORD, 2015).

Em *H. salinarum* NRC-1, encontramos 16 asRNAs sobrepostos a região 5' de seus genes que codificam transposases, nenhum deles apresenta potencial interação com a proteína LSm. Esses casos podem constituir exemplos de reguladores que agem de forma similar ao art200. Tal prevalência de asRNAs localizados em transposases em várias arqueias é um indício de que, similar ao que já é conhecido em bactérias, esses ncRNAs possuem um papel funcional na regulação da transposição em Archaea.

5.6.3-asRNAs nos genes codificadores de Rodopsinas

A proteína de membrana bacteriorodopsina confere a *H. salinarum* NRC-1 duas de suas características marcantes, a coloração púrpura característica e a capacidade de produzir

ATP utilizando luz (OESTERHELT; STOECKENIUS, 1971). Essa proteína é codificada pelo gene *bop* (VNG_RS05715), encontrado em cópia única no genoma de *H. salinarum* NRC-1.

Nesse estudo identificamos um asRNA presente nesse gene, o que pode indicar um importante papel desse ncRNA como seu regulador transcricional. Além desse, também identificamos um asRNAs sobreposto ao gene *brb* que está associado a regulação do gene *bop*. Devido a sua importância fisiológica para esse organismo, o gene *bop* possui seu produto proteico e seus reguladores transcricionais amplamente caracterizados (BALIGA et al., 2001). Tarasov et al. (2011) foram os primeiros a caracterizar o gene *brb*, que possui papel na regulação fina da ativação do gene *bop*. Tal função foi evidenciada com experimentos utilizando repórteres e mutagênese, porém a proteína Brb nunca foi detectada (TARASOV et al., 2011). Nas bibliotecas de RNA-seq analisadas nesse estudo, a média da contagem de reads entre as réplicas para esse asRNA é sempre maior que a para o gene *brb*, em média quatro vezes maior. A presença de um asRNA sobrepondo o gene *brb* pode indicar que o pareamento entre o asRNA e seu mRNA pode levar a degradação ou impedimento da tradução do mesmo, o que explicaria o fenômeno observado por Tarasov et al. (2011), caracterizando mais um elemento regulador pós-transcricional relacionado a bacteriorodopsina.

Também identificamos um asRNA para o gene VNG_RS00745 que codifica para a proteína halorodopsina. Essa proteína possui um grupo retinal que sofre modificações quando interage com a luz e atua como uma bomba de íons pelos mesmos princípios que a bacteriorodopsina, porém se distingue dessa última por bombear íons de cloreto através da membrana (OESTERHELT; TITTOR, 1989).

Apresentamos evidências de que asRNAs podem estar envolvidos na regulação de duas proteínas de extrema importância biológica para esse organismo e que também foram alvo de estudos no campo da fotobiologia que ajudaram a estabelecer esse organismo como um importante modelo de estudos do domínio Archaea.

5.6.4-asRNAs em genes relacionados a vesículas de gás – (*gas vescicle proteins* – *gvp*)

A produção de vesículas de gás de natureza proteica em *H. salinarum* NRC-1 é uma característica icônica dessa arqueia. O processo de formação dessas vesículas é complexo, envolve 14 diferentes genes *gvp*. Esses genes estão agrupados em regiões genômicas denominadas *vac regions* (Figura 6) (PFEIFER et al., 1997). Estruturalmente, as vesículas de

gás são constituídas pelos produtos proteicos dos genes *gvpA* e *gvpC*. Os demais genes codificam proteínas estruturais complementares ou reguladoras da expressão gênica dos demais genes de uma *vac region* (PFEIFER, 2012). Uma dessas proteínas reguladoras é codificada pelo gene *gvpD*. Kruger e Pfeifer (1996) encontraram dois asRNAs complementares respectivamente as porções 5' e 3' desse gene em *H. salinarum*. A detecção foi realizado por Northern blot, a mesma técnica foi utilizada pelos autores para mostrar que esses asRNAs são detectáveis em momentos do crescimento em que a proteína codificada por esse gene apresenta baixos níveis.

O asRNA (164 nt) que identificamos na extremidade 5' do gene *gvpD*, possui tamanho similar ao da banda de 190 nt identificada por Kruger e Pfeifer (1996) para a mesma região antisense desse gene. Além disso, a posição em que o aTSS desse asRNA foi encontrado fica a apenas 3 nt *downstream* ao início da sonda utilizada. Evidenciando que por técnicas diferentes, ambos resultados podem se tratar do mesmo asRNA sobreposto a extremidade 5' de *gvpD*.

Também identificamos asRNA com aTSSs para os genes, *gvpCHIJ*, e para o gene *gvpA* encontramos um nítido sinal de RNA-seq antisense a extremidade 5', a ausência de um TSS para esse asRNA indica que o mesmo pode ser resultado de algum tipo de processamento.

asRNAs como possíveis reguladores da formação de vesículas de gás não é um fenômeno que se restringe a *H. salinarum*. Para a cianobactéria *Calothrix* sp. PCC 7601, que possui uma *vac region* constituída de apenas quatro genes, um asRNA de 420nt foi encontrado. Esse asRNA sobrepõe dois genes *gvp* desse organismo e sua expressão pode estar relacionada a regulação da formação de vesículas de gás em momentos de diferenciação celular desse organismo (CSISZAR; HOUMARD; DAMERVAL, TANDEAU DE MARSAC, 1987).

Apresentamos evidências da presença de asRNAs para os dois genes que codificam os principais constituintes estruturais das vesículas de gás em *H. salinarum* NRC-1 (*gvpA* e *C*), além de recuperarmos em nossos resultados um asRNA já conhecido para o gene codificador da proteína reguladora GvpD. Também encontramos asRNAs para outros quatro genes (*gvpHIJL*), constituinte da *vac region* *gvp-1*. Nossos resultados evidenciam que a regulação

da formação de vesículas de gás por asRNAs pode ser ainda mais relevante do que já era predito em *H. salinarum*.

5.7-asRNAs em TAs tipo II em bactérias e arqueias

Em bactérias, sistemas toxinas-antitoxinas (TAs) do Tipo I são conhecidos pela presença de um asRNA atuando como antitoxina (Figura 11B), enquanto nos TAs do tipo II, ambas são proteínas (Figura 12). Lybecker et al. (2014), realizando imunoprecipitações utilizando anticorpos específicos para dsRNAs procedidas de RNA-seq, encontraram asRNAs sobrepostos a dois genes codificadores de toxinas em TAs do tipo II. A presença desses asRNAs também foi confirmada por *Northern Blot* no mesmo estudo. Tal resultado foi surpreendente, pois há muito tempo estava estabelecido que apenas TAs do tipo I possuíam asRNAs e tal descoberta evidência que a regulação de TAs do tipo II pode ser mais complexa do que se esperava, envolvendo não apenas duas proteínas, mas também asRNAs. Ao que nossa revisão de literatura indica, a presença de asRNAs em TAs do tipo II não recebeu atenção desde a publicação de Lybecker et al. (2014), nem mesmo nas revisões mais recentes sobre TAs (CORAY et al., 2017; HARMS et al., 2018).

Nesse estudo, utilizamos a ferramenta *TA finder* 2.0 para a anotação de TAs do tipo II nos genomas das arqueias que possuem dados de dRNA-seq disponíveis para verificar se, com essa técnica, conseguimos identificar asRNAs nos pares de genes componentes desses sistemas nesses organismos e em *E. coli*. A maioria dos TAs do tipo II identificados nos genomas das arqueias ainda não estavam anotados, ou ao menos não constavam nas anotações disponíveis para os genomas no NCBI, o que pode ser um reflexo da pouca quantidade de estudos sobre TAs no domínio Archaea. Encontramos aTSSs para a maioria dos TAs do tipo II anotados nos genomas de todos organismos analisados. Para os archeons *T. kodakarensis* e *M. jannaschii* e para *E. coli*, a presença de aTSSs é predominante nas toxinas, porém não fica nítido se um possível modo de ação de asRNAs nesses sistemas seria reprimindo a expressão das toxinas assim como em TAs do tipo I, pois também encontramos uma grande quantidade de genes de antitoxinas com aTSSs em todos genomas analisados.

Em *H. pylori* e *B. Subtilis*, dados de dRNA-seq permitiram a identificação de sistemas de asRNAs em TAs do tipo I (SHARMA et al., 2010) (IRNOV et al., 2010), evidenciando a utilidade dessa técnica no estudo de TAs. Nossos resultados evidenciam que asRNAs podem

estar presentes em vários genes de TAs do tipo II, tanto em toxinas, quanto em antitoxinas, em arqueias e bactérias. Mais estudos são necessários para verificar possíveis mecanismos e impactos funcionais dessas moléculas na regulação desses TAs. Para alguns casos, é possível que estamos observando um novo tipo de TAs, ou ao menos, um subtipo de TAs do tipo II.

5.8-Expressão diferencial na curva de crescimento

Dentre os genes diferencialmente expressos nas duas transições, não encontramos pares asRNA/gene com genes cujo aumento ou diminuição da expressão são característicos do estado fisiológico daquele momento da curva de crescimento em *H. salinarum* NRC-1 descritos por Faccioli et al. (2010). Na transição para a fase exponencial, dois genes que codificam proteínas *cdc6*, VNG_RS00175 ($\log_2FC = -1.3$) e VNG_RS04790 ($\log_2FC = -1.6$) estão *downregulados*. Essas proteínas são fundamentais para o início da replicação em eucariotos e na arqueia, *Pyrococcus abyssi*, são encontradas em grande quantidade interagindo com o DNA durante a fase exponencial de crescimento (MATSUNAGA et al., 2001). A observação dos genes codificadores dessas proteínas *downregulados* no início da fase exponencial vai na direção contrária da inferência de que, por estarem envolvidas na replicação do material genético, seriam encontrados *upregulados* nesse momento. Os níveis de expressão dos asRNAs desses genes também dificultam a interpretação de um possível envolvimento nesse fenômeno, pois para um dos genes está *upregulado* e para o outro *downregulado*. Porém, na lista de genes que não possuem asRNAs anotados e que se encontram *upregulados* nessa transição, encontramos os genes VNG_RS08860 ($\log_2FC = 1.11$) e VNG_RS11810 ($\log_2FC = 1.44$), ambos codificadoras de proteínas *cdc6*. Uma possível explicação, é que os asRNAs podem estar envolvidos em um mecanismo de controle de quais *cdc6* serão expressas nessa fase.

Também na primeira transição, encontramos diferencialmente expresso um par asRNA/gene contendo um gene codificador de uma toxina do tipo II. O gene VNG_RS00140 ($\log_2FC = 2.7$) está *upregulado*, enquanto seu asRNA ($\log_2FC = -1.4$) é encontrado *downregulado*. A toxina codificada por esse gene está envolvida na inibição da tradução, sendo um fenômeno intrigante o aumento da expressão desse gene no início da fase exponencial (JØRGENSEN et al., 2009). A inibição das toxinas em TAs do tipo II ocorre por interação proteína-proteína entre a toxina com a antitoxina (HARMS et al., 2018) (Figura 12),

logo, esperávamos encontrar o gene VNG_RS00135, codificador da antitoxina, também *upregulado*, o que não ocorreu. A relação inversa dos níveis de expressão do asRNA com a toxina, é um fenômeno observado em TAs do tipo I, onde o asRNA é a antitoxina (HAYES; VAN MELDEREN, 2011). Uma diferença importante é a de que, para o gene VNG_RS00140 o asRNA não sobrepõe a extremidade 5' gene, detalhe a ser mencionado já que essa sobreposição em TAs do tipo I determina o mecanismo de ação do asRNA como inibidor da interação do mRNA da toxina com a maquinaria ribossomal (Figura 11D). Para o asRNA sobrepondo o gene VNG_RS11890 que codifica uma toxina do tipo do II (Figura 31) e para os asRNAs encontrados por Lybecker et al. (2014), o mesmo perfil de não sobreposição da extremidade 5' do gene codificador de toxinas por asRNAs em TAs do tipo II é encontrado. A observação desse par asRNA/toxina diferencialmente expressos com relação inversa, reforça nossa hipótese de que asRNAs em TAs do tipo II podem se tratar de importantes constituintes regulatórios desses sistemas.

Apenas 15 asRNAs diferencialmente expressos e interagindo com LSm, também possuem o gene sense e com transcritos também interagindo com essa proteína (Figura 35). A baixa quantidade de potenciais reguladores em *cis* interagindo com LSm é um resultado que vai de encontro a hipótese de que esse tipo de regulação não necessita uma proteína mediadora por conta da proximidade e complementariedade completa de pares de base entre o asRNA e o mRNA (VOGEL, JÖRG; LUISI, 2011). Além disso, 13 dos pares interagindo com LSm são de genes que codificam proteínas hipotéticas e em todos esses pares, ambos asRNA e gene se encontram *upregulados*. Os genes codificadores dessas proteínas compõe 27 dos 56 pares diferencialmente expressos (Figura 35). Um estudo recente que avaliou a expressão diferencial de sRNAs sob condições de estresse oxidativo em *H. volcanii*, também encontrou uma grande quantidade (48%) de genes que codificam proteínas hipotéticas diferencialmente expressos quando seus asRNAs também estavam. Esses resultados indicam uma provável regulação desses genes por asRNAs em ambos halófilos, porém o impacto funcional da expressão desses genes para ambos organismos permanece desconhecido.

Dentre os 56 pares diferencialmente expressos, em 14 desses, o asRNA pode, de alguma forma, estarem relacionados a interações da maquinaria ribossomal com o mRNA de seus genes sense por serem encontrados sobrepondo a extremidade 5' desses (Figura 35).

Análises de expressão diferencial de dados de transcriptoma são comumente utilizadas para relacionar conjuntos de genes diferencialmente expressos a fenótipos observados em diferentes tratamentos ou situações a quais um organismo é exposto. Porém, apesar dos níveis de expressão de um gene serem indicadores de sua relevância funcional para um momento ou estado de um organismo, nem sempre a relação entre níveis de expressão e tradução é positiva (VOGEL, CHRISTINE; MARCOTTE, 2013). Em muitos casos, essa falta de correlação é atribuída a mecanismos de regulação pós-transcricional, como por exemplo, os envolvendo ncRNAs. Para os asRNAs, é esperado que a alteração dos níveis dessas moléculas, quando acompanhada de mudanças nos níveis de expressão do gene sense, altere a quantidade de proteínas produzida. Seguindo esse pressuposto, muitas inferências sobre possíveis mecanismos de regulação são realizadas analisando os níveis de expressão dos pares asRNA/gene em dados de larga escala. Alguns autores alertam que é necessário cautela para inferir regulação a partir de dados de expressão para asRNAs. A regulação da expressão do gene sense e do antisense pode ser independente, de modo que a interferência na expressão de um não altera o nível do outro (GOYAL et al., 2017). Utilizando construções gênicas e ensaios de fluorescência com a proteína GFP, Huber et al. (2016) avaliaram os níveis de tradução e da expressão antisense de 188 genes de *S. cerevisiae*. Os autores evidenciaram que em apenas 25% dos genes a expressão antisense afeta a produção de proteínas (HUBER et al., 2016). Em nossas análises de expressão diferencial, observamos que dos 144 asRNAs diferencialmente expressos, apenas 56 possuem a expressão do gene sense também alterada e a maioria desses pares possui sua expressão positivamente correlacionada. Como já discutido, é difícil inferir se esses 56 asRNAs estão realmente atuando em *cis* e se os 88 asRNAs restantes são potenciais reguladores em *trans*. A lista de asRNAs diferencialmente expressos gerada nesse estudo pode servir como ponto de partida para experimentos ou análises mais detalhadas que poderão fornecer novas evidências sobre o papel regulador dessas moléculas.

5.9-aTSSs conservados em *H. volcanii*

O organismo filogeneticamente mais próximo de *H. salinarum* NRC-1 com dados de dRNA-seq disponíveis para comparação é a arqueia *H. volcanii*. Em nossa análise comparativa, verificamos que 18% dos asRNAs anotados em *H. salinarum* NRC-1 possui um potencial aTSS conservado em *H. volcanii*.

Buscando por aTSSs que se encontram exatamente na mesma posição em genes ortólogos nos dois organismos, encontramos dois casos (Figura 36). Baseado na anotação dos asRNAs em *H. salinarum* NRC-1 as duas sequências correspondentes aos asRNAs desses aTSSs apresentam alta similaridade entre os dois organismos. Porém, em análises da conservação de sequências de asRNAs para inferência de homologia, sempre permanece a incerteza se a porcentagem de identidade observada não se trata apenas de um resultado de pressão seletiva sobre a sequência codificadora do gene. Apesar disso, a conservação exata do início de transcrição é uma forte evidência de conservação desses asRNAs entre esses dois organismos.

Várias análises de transcritomas de bactérias mostram uma baixa conservação de asRNAs, mesmo em espécies filogeneticamente próximas. A comparação de asRNAs entre as bactérias entéricas *E. coli* e *Salmonella enterica* serovar Typhimurium mostrou que apenas 14% de asRNAs inferidos a partir de dados de RNA-seq são conservados entre essas duas espécies (BREU; GUGGENBICHLER; WOLLMANN, 2008). A comparação entre quatro linhagens de bactérias do gênero *Campylobacter* resultou em 45% dos aTSSs identificados por dRNA-seq conservados (DUGAR et al., 2013). Em um estudo avaliando a conservação de TSSs em oito bactérias do gênero *Shewanella*, os aTSSs constituíram a classe menos conservada entre os organismos, com apenas 22% conservados (SHAO et al., 2014). Entre as linhagens de cianobactéria *Synechocystis* sp. PCC 6714 e PCC 6803, apenas as 4% dos aTSS identificados por dRNA-seq foram considerados conservados (KOPF et al., 2015).

Um pressuposto básico para inferir funcionalidade a novos genes é o de que ao longo do processo evolutivo, genes funcionais terão suas sequências conservadas entre espécies. A falta de conservação entre asRNAs de espécies filogeneticamente próximas, assim como seus baixos níveis de expressão, levam ao questionamento sobre o papel funcional de boa parte dos asRNAs identificados por métodos de larga escala (SHAO et al., 2014). Há evidências de que as altas taxas de evolução de ncRNAs estariam relacionadas há uma adaptação mais rápida dos organismos em relação ao remodelamento de vias regulatórias para gerar respostas celulares específicas, o que se aplicaria aos asRNAs (KOPF et al., 2015; PELECHANO; STEINMETZ, 2013).

Nossa comparação mostra uma pequena quantidade de aTSSs potencialmente conservados e apenas dois aTSS com a posição exata conservada entre genes irtólogos. Esses

resultados indicam que muitos dos asRNAs encontrados em *H. salinarum*, podem se tratar de possíveis elementos reguladores específico desse organismo ou de espécies com distâncias filogenéticas menores que a existente entre *H. salinarum* e *H. volcanii*.

5.10-RNase R

RNases R são enzimas versáteis, podendo atuar como exoribonucleases, degradando RNAs fita simples na direção 3'-5' (VINCENT; DEUTSCHER, 2009) e também podem degradar RNAs estruturados e dupla fita em bactérias (CHENG; DEUTSCHER, 2005; MATOS, RUTE; BARBAS; ARRAIANO, 2009). Postula-se que sejam essenciais para muitas arqueias por se tratar da única exoribonuclease que codificam em seus genomas (PORTNOY; SCHUSTER, 2006). Essa enzima purificada do halófilo *H. volcanii* apresentou atividade de degradação de dsRNAs *in vitro* (MATOS, RUTE et al., 2012).

Em relação as tentativas de obter linhagens *knockout* para o gene dessa RNase em *H. volcanii*, a simples obtenção da linhagem *knockout* viável em *H. salinarum* NRC-1 já foi um resultado surpreendente, pois contraria as expectativas de que essa RNase seria essencial para o crescimento de muitas arqueias por se tratar da única exoribonuclease codificada no genoma de muitos desses organismos (PORTNOY; SCHUSTER, 2006). Uma das implicações desse resultado, é que as vias essenciais de degradação de RNAs em *H. salinarum* NRC-1 provavelmente ocorrem por intermédio de outras enzimas, cujas atividades exoribocatalíticas permanecem desconhecidas.

Em bactérias, a RNase R está envolvida em processos de degradação e modificação funcional de tRNAs (LI, ZHONGWEI et al., 2002; WELLNER et al., 2017), porém não há relatos de envolvimento de asRNAs em nenhum desses. Dentre os demais genes *upregulados* que não possuem asRNAs anotados, encontramos mais dois genes que codificam tRNAs, o gene VNG_RS05810 (log2FC = 1.45) do tRNA-Leu e o gene VNG_RS06715 (log2FC = 1.77) do tRNA-Pro. tRNAs são RNAs altamente estruturados e passam por processos pós-transcricionais mediados por RNases (SCHIERLING et al., 2002). Encontrar genes referentes a esses ncRNAs *upregulados* no *knockout* de uma RNase R é algo esperado por se tratar de uma RNase com que pode degradar RNAs estruturados. Porém, não parece haver relatos de processamento de tRNAs mediado por asRNAs na literatura. O possível duplex asRNA/RNA que relatamos como *upregulado* nesse estudo pode se tratar de um alvo específico dessa

RNase em *H. salinarum* NRC-1, constituindo algum mecanismo de controle de expressão do tRNA-Lys ainda desconhecido.

Na lista de genes *upregulados* da linhagem *knockout*, também encontramos uma possível explicação para o fenótipo observado a partir das 50 horas na curva de crescimento (Figura 39). Após a primeira fase exponencial de crescimento de *H. salinarum* NRC-1, que acaba por volta das 50 horas em uma curva de crescimento, o aumento dos valores de OD600 podem não estar relacionados ao aumento de células, mas sim a produção de pigmentos e liberação de vesículas de gás (FACCIOTTI et al., 2010). O gene *bop* (VNG_RS05715) que codifica a bacteriorodopsina, está *upregulado* ($\log_2FC = 3.23$) na linhagem *knockout*. Não observamos alterações na expressão do asRNA presente nesse gene, o que descarta a hipótese de uma possível regulação envolvendo o duplex asRNA/mRNA e a RNase R. De qualquer forma, tal resultado pode estar relacionado ao fenótipo que observamos na curva de crescimento que provavelmente não se trata de aumento do crescimento celular na ausência dessa RNase, mas sim no envolvimento da mesma na regulação de um importante gene para *H. salinarum* NRC-1.

Nossos resultados não suportam a hipótese de que a RNase R desempenha um papel fundamental na degradação de dsRNAs em *H. salinarum* NRC-1, ou não se trata de um fenômeno detectável por uma abordagem em larga escala como a utilizada nesse estudo. O fato dessa enzima purificada de *H. volcanii* ter apresentado tal atividade catalítica *in vitro*, não garante que desempenha a mesma função no ambiente citoplasmático, questão que continuará em aberto por conta das tentativas falhas de obtenção de uma linhagem *knockout* do seu gene nesse organismo.

Elucidar os mecanismos de degradação de dsRNAs é sem dúvidas um dos pontos-chaves para aumentar nossa compreensão sobre a atuação de asRNAs no terceiro domínio da vida. Essas moléculas são ubíquas nos transcritomas das arqueias até hoje avaliadas. A descoberta de uma RNase alternativa ou de mecanismos de regulação pós-transcricionais ainda desconhecidos envolvendo essas moléculas no domínio Archaea será um grande avanço não só para os estudos dos asRNAs, mas também para o campo de estudos dos ncRNAs.

6-Conclusões

Esse estudo fornece um mapa abrangente do transcritoma antisense primário de *H. salinarum* NRC-1. Os resultados apresentados evidenciam que, assim como em outros organismos pertencentes aos três domínios da vida, a transcrição antisense é um fenômeno pervasivo em *H. salinarum* NRC-1, gerando asRNAs que possivelmente estão envolvidos na regulação pós-transcricional de genes codificadores de proteínas relacionadas a características marcantes desse organismo, como as rodopsinas e vesículas de gás. Pela pequena quantidade de pares asRNA/mRNA encontrados interagindo concomitante com a proteína LSM, não podemos inferir que essa proteína se trata de uma mediadora ubíqua da formação desses duplex em *H. salinarum* NRC-1. Nossas evidências vão de encontro a hipótese de que para esse tipo de regulação envolvendo asRNAs, a presença de uma proteína mediadora é de menor relevância, ou, outras proteínas desconhecidas estão envolvidas. Além de encontrarmos aTSSs e asRNAs em TAs do tipo II em *H. salinarum* NRC-1, encontramos aTSSs na maioria desses sistemas em arqueias que possuem dados de dRNAseq, evidenciando que o modo de regulação desses genes pode envolver asRNAs em bactérias e arqueias. Encontramos asRNAs com potencial papel de reguladores em *cis* ao longo da curva de crescimento, esses podem se tratar de ncRNAs importantes nas transições de estado fisiológico desse organismo, consistindo em alvos para futuros estudos do fenômeno de regulação pós-transcricional antisense em *H. salinarum* NRC-1. Tal como relatado para outros organismos, apenas uma pequena quantidade de aTSSs são potencialmente conservados entre *H. salinarum* e a arqueia filogeneticamente próxima, *H. volcanii*. A ausência de dados de dRNA-seq para organismos filogeneticamente mais próximos (e.g. do mesmo gênero) nos impede avaliarmos se a maioria dos asRNAs encontrados nesse estudo se tratam de possíveis reguladores pós-transcricionais exclusivos de *H. salinarum* NRC-1. Nossa abordagem em larga escala para verificar a influência da deleção do gene da RNase R no transcritoma antisense primário de *H. salinarum* NRC-1 e seu possível papel na degradação de dsRNAs não nos forneceu evidências de que essa RNase tem relação com esses dois fenômenos nesse organismo. Os resultados aqui apresentados consistem em um importante passo na direção da compreensão do envolvimento da transcrição antisense na regulação gênica pós-transcricional de *H. salinarum* NRC-1 e conseqüentemente, na compreensão desse fenômeno no terceiro domínio da vida.

7-Referências

- ALLERS, Thorsten; MEVARECH, Moshe. Archaeal genetics - the third way. *Nature reviews. Genetics*, v. 6, n. January, p. 58–73, 2005. Disponível em: <<http://www.nature.com.ezproxy.weizmann.ac.il/nrg/journal/v6/n1/full/nrg1504.html>>.
- ALTSCHUL, S F et al. Basic local alignment search tool. *Journal of molecular biology*, v. 215, n. 3, p. 403–10, 1990. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283605803602>>.
- AMMAN, Fabian et al. TSSAR: TSS annotation regime for dRNA-seq data. *BMC Bioinformatics*, v. 15, n. 1, 2014.
- ANDERS, Simon; PYL, Paul Theodor; HUBER, Wolfgang. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, v. 31, n. 2, p. 166–169, 2015.
- ANGRAND, Pierre Olivier et al. The role of long non-coding RNAs in genome formatting and expression. *Frontiers in Genetics*, v. 6, n. APR, p. 1–12, 2015.
- AOKI-KINOSHITA, Kiyoko F.; KANEHISA, Minoru. Gene annotation and pathway mapping in KEGG. *Methods in Molecular Biology*, v. 396, p. 71–91, 2007.
- BABSKI, Julia et al. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genomics*, v. 17, n. 1, p. 629, 2016. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/27519343>%5Cn<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4983044>%5Cn<http://bmcbgenomics.biomedcentral.com/articles/10.1186/s12864-016-2920-y>%5Cn<http://www.ncbi.nlm.nih.gov/pubmed/27519343>%5Cn<http://www.pub>>.
- BABSKI, Julia et al. Small regulatory RNAs in Archaea. *RNA biology*, NULL, v. 11, n. 5, p. 484–93, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4152357&tool=pmcentrez&rendertype=abstract>>.
- BACHELLERIE, Jean Pierre; CAVAILLÉ, Jérôme; HÜTTENHOFER, Alexander. The expanding snoRNA world. *Biochimie*, v. 84, n. 8, p. 775–790, 2002.
- BALAKRISHNAN, Arjun et al. Halobacterial nano vesicles displaying murine bactericidal permeability-increasing protein rescue mice from lethal endotoxic shock. *Scientific reports*, NULL, v. 6, n. August, p. 33679, 2016. Disponível em: <<http://dx.doi.org/10.1038/srep33679>>.
- BALIGA, N. S. et al. Genomic and genetic dissection of an archaeal regulon. *Proceedings of the National Academy of Sciences*, v. 98, n. 5, p. 2521–2525, 2001. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.051632498>>.
- BARE, J. Christopher et al. Integration and visualization of systems biology data in context of the genome. *BMC Bioinformatics*, v. 11, 2010.
- BEER, Karlyn D. et al. Model organisms retain an “ecological memory” of complex ecologically relevant environmental variation. *Applied and Environmental Microbiology*, v. 80, n. 6, p. 1821–1831, 2014.

- BEITER, T. et al. Antisense transcription: A critical look in both directions. *Cellular and Molecular Life Sciences*, v. 66, n. 1, p. 94–112, 2009.
- BELL, Stephen D.; JACKSON, Stephen P. Mechanism and regulation of transcription in archaea. *Current Opinion in Microbiology*, v. 4, n. 2, p. 208–213, 2001.
- BERNSTEIN, H D et al. Functional substitution of the signal recognition particle 54-kDa subunit by its Escherichia coli homolog. *Proc Natl Acad Sci U S A*, v. 90, n. 11, p. 5229–5233, 1993.
- BILES, Benjamin D.; CONNOLLY, Bernard A. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. *Nucleic acids research*, v. 32, n. 22, 2004.
- BISCHLER, Thorsten et al. Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in Helicobacter pylori. *Methods*, v. 86, 2015.
- BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, v. 30, n. 15, p. 2114–2120, 2014.
- BONNEAU, Richard et al. A Predictive Model for Transcriptional Control of Physiology in a Free Living Cell. *Cell*, v. 131, n. 7, p. 1354–1365, 2007.
- BOUVIER, Marie et al. Small RNA Binding to 5' mRNA Coding Region Inhibits Translational Initiation. *Molecular Cell*, v. 32, p. 827–837, 2008.
- BOVRE, K; SZYBALSKI, W. Patterns of Convergent the Overlapping of Coliphage Transcription A " ~ within b2 Region Most of the functions and corresponding mined and mapped (see review by Dove , and later identified as a deletion by the *et al .*, 1960) and in the sedimentation co-. *Virology*, NULL, v. 38, p. 614–626, 1969.
- BRENES-ÁLVAREZ, Manuel et al. Identification of conserved and potentially regulatory small RNAs in heterocystous cyanobacteria. *Frontiers in Microbiology*, v. 7, n. FEB, p. 1–11, 2016.
- BRENNEIS, Mariam et al. Experimental characterization of cis-acting elements important for translation and transcription in halophilic archaea. *PLoS Genetics*, v. 3, n. 12, p. 2450–2467, 2007.
- BRENNER, S.; JACOB, F.; MESELSON, M. An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, v. 190, n. 4776, p. 576–581, 1961.
- CARNINCI, Piero et al. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, v. 38, n. 6, p. 626–635, 2006.
- CARRIERI, Claudia et al. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature*, v. 491, n. 7424, p. 454–457, 2012. Disponível em: <<http://dx.doi.org/10.1038/nature11508>>.
- CAVICCHIOLI, Ricardo. Archaea - Timeline of the third domain. *Nature Reviews Microbiology*, v. 9, n. 1, p. 51–61, 2011. Disponível em: <<http://dx.doi.org/10.1038/nrmicro2482>>.

- CHABAN, Bonnie; NG, Sandy Y.M; JARRELL, Ken F. Archaeal habitats — from the extreme to the ordinary. *Canadian Journal of Microbiology*, v. 52, n. 2, p. 73–116, 2006. Disponível em: <<http://www.nrcresearchpress.com/doi/abs/10.1139/w05-147>>.
- CHEN, Shuo et al. MicC , a Second Small-RNA Regulator of Omp Protein Expression in Escherichia coli. *Journal of bacteriology*, v. 186, n. 20, p. 6689–6697, 2004.
- CHENG, Zhuan Fen; DEUTSCHER, Murray P. An important role for RNase R in mRNA decay. *Molecular Cell*, v. 17, n. 2, p. 313–318, 2005.
- CHO, Suhyung et al. Genome-wide primary transcriptome analysis of H₂-producing archaeon *Thermococcus onnurineus* NA1. *Scientific Reports*, v. 7, n. February, p. 43044, 2017. Disponível em: <<http://www.nature.com/articles/srep43044>>.
- COKER, James a et al. Transcriptional profiling of the model Archaeon *Halobacterium* sp. NRC-1: responses to changes in salinity and temperature. *Saline systems*, v. 3, n. February 2007, p. 6, 2007.
- CONESA, Ana et al. A survey of best practices for RNA-seq data analysis. *Genome Biology*, v. 17, n. 1, p. 1–19, 2016.
- CORAY, Dorien S. et al. Why so narrow: Distribution of anti-sense regulated, type I toxin-antitoxin systems compared with type II and type III systems. *RNA Biology*, v. 14, n. 3, p. 275–280, 2017. Disponível em: <<http://dx.doi.org/10.1080/15476286.2016.1272747>>.
- CRICK, F H. On protein synthesis. *Symposia of the Society for Experimental Biology*, v. 12, p. 138–163, 1958.
- CROOKS, G et al. NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Research*, v. 14, p. 1188–1190, 2004. Disponível em: <<ftp://ftp.ncbi.nih.gov/genomes/Bacteria>>.
- CSISZAR, K; HOUMARD, J; DAMERVAL, T.& TANDEAU DE MARSAC, N. Transcriptional analysis of the cyanobacterial gvgABC operon in differentiated cells: occurrence of an antisense RNA complementary to three overlapping transcripts. *Gene*, v. 60, p. 29–37, 1987.
- DAR, Daniel et al. Widespread formation of alternative 3' UTR isoforms via transcription termination in archaea. *Nature Microbiology*, v. 1, n. 10, p. 1–9, 2016. Disponível em: <<http://dx.doi.org/10.1038/nmicrobiol.2016.143>>.
- DASSARMA, Shiladitya. *Extreme Microbe*. v. 51, n. 2, p. 180–192, 2015.
- DECATUR, Wayne A. et al. Identifying Effects of snoRNA-Guided Modifications on the Synthesis and Function of the Yeast Ribosome. *Methods in Enzymology*. [S.l: s.n.], 2007. v. 425. p. 283–316.
- DENNIS, Patrick P.; OMER, Arina. Small non-coding RNAs in Archaea. *Current Opinion in Microbiology*, v. 8, n. 6, p. 685–694, 2005.

DOENCH, John G. Am I ready for CRISPR? A user's guide to genetic screens. *Nature Reviews Genetics*, v. 19, n. 2, p. 67–80, 2017. Disponível em: <<http://www.nature.com/doi/10.1038/nrg.2017.97>>.

DOLGIN, Elie. Phylogeny: Rewriting evolution. *Nature*, 2012. , v. 486, n. 7404, p. 460–462.

DYALL-SMITH, M. The Halohandbook: protocols for halobacterial genetics. *Mark Dyall-Smith, Martinsried, Germany*, n. March, p. 1–144, 2008. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Halohandbook#0>>.

ECKWEILER, Denitsa; HÄUSSLER, Susanne. Antisense transcription in *Pseudomonas aeruginosa*. *Microbiology*, n. May, p. 1–7, 2018. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/29738307>><http://www.microbiologyresearch.org/content/journal/micro/10.1099/mic.0.000664.v1>>.

EDDY, Sean R. Non-coding RNA genes and the modern RNA world. *Nature Reviews Genetics*, v. 2, n. 12, p. 919–929, 2001.

EHLERS, Claudia; JÄGER, Dominik; SCHMITZ, Ruth A. Establishing a markerless genetic exchange system for *Methanosarcina mazei* strain G??1 for constructing chromosomal mutants of small RNA genes. *Archaea*, v. 2011, 2011.

ELLIS, Michael J.; TRUSSLER, Ryan S.; HANIFORD, David B. A cis-encoded sRNA, Hfq and mRNA secondary structure act independently to suppress IS200 transposition. *Nucleic Acids Research*, v. 43, n. 13, p. 6511–6527, 2015.

EME, Laura et al. Archaea and the origin of eukaryotes. *Nature Reviews Microbiology*, v. 15, n. 12, p. 711–723, 2017.

FACCIOTTI, Marc T et al. General transcription factor specified global gene regulation in archaea. *Proceedings of the National Academy of Sciences of the United States of America*, v. 104, p. 4630–4635, 2007.

FACCIOTTI, Marc T. et al. Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. *BMC Systems Biology*, v. 4, p. 1–19, 2010.

FAGHIHI, Mohammad et al. Evidence for natural antisense transcript-mediated \ninhibition of microRNA function. p. 1–13, 2010.

FILIPOWICZ, Witold et al. Post-transcriptional gene silencing by siRNAs and miRNAs. *Current Opinion in Structural Biology*, v. 15, n. 3 SPEC. ISS., p. 331–341, 2005.

FISCHER, Susan et al. The archaeal lsm protein binds to small RNAs. *Journal of Biological Chemistry*, v. 285, n. 45, p. 34429–34438, 2010.

FOZO, E. M.; HEMM, M. R.; STORZ, G. Small Toxic Proteins and the Antisense RNAs That Repress Them. *Microbiology and Molecular Biology Reviews*, v. 72, n. 4, p. 579–589, 2008. Disponível em: <<http://mmb.asm.org/cgi/doi/10.1128/MMBR.00025-08>>.

FRIEDMAN, Robin C et al. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, p. 92–105, 2009.

GARALDE, Daniel R et al. Highly parallel direct RNA sequencing on an array of nanopores. *Nature Methods*, n. November 2017, 2018. Disponível em: <<http://www.nature.com/doi/10.1038/nmeth.4577>>.

GARBER, Manuel et al. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature Methods*, v. 8, n. 6, p. 469–477, 2011. Disponível em: <<http://dx.doi.org/10.1038/nmeth.1613>>.

GE, Hui; WALHOUT, Albertha J.M.; VIDAL, Marc. Integrating “omic” information: A bridge between genomics and systems biology. *Trends in Genetics*, v. 19, n. 10, p. 551–560, 2003.

GELSINGER, Diego Rivera; DIRUGGIERO, Jocelyne. The non-coding regulatory RNA revolution in archaea. *Genes*, v. 9, n. 3, 2018a.

GELSINGER, Diego Rivera; DIRUGGIERO, Jocelyne. Transcriptional Landscape and Regulatory Roles of Small Noncoding RNAs in the Oxidative Stress Response of the Haloarchaeon. *JOURNAL OF BACTERIOLOGY*, v. 200, n. 9, p. 1–17, 2018b.

GEORG, Jens; HESS, Wolfgang R. cis-antisense RNA, another level of gene regulation in bacteria. *Microbiology and molecular biology reviews : MMBR*, NULL, v. 75, n. 2, p. 286–300, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3122628&tool=pmcentrez&rendertype=abstract>>.

GILBERT, Walter. The RNA world. *Nature*, v. 319, n. February, p. 618, 1986.

GOMES-FILHO, José Vicente et al. Sense overlapping transcripts in IS1341-type transposase genes are functional non-coding RNAs in archaea. *RNA biology*, v. 12, n. 5, p. 490–500, 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/25806405>>.

GOTTESMAN, Susan; STORZ, Gisela. Bacterial small RNA regulators: Versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology*, v. 3, n. 12, p. 1–16, 2011.

GOYAL, Ashish et al. A cautionary tale of sense-Antisense gene pairs: Independent regulation despite inverse correlation of expression. *Nucleic Acids Research*, v. 45, n. 21, p. 12496–12508, 2017.

GRABOWSKI, Beatrice; KELMAN, Zvi. Archaeal DNA Replication: Eukaryal Proteins in a Bacterial Context. *Annual Review of Microbiology*, v. 57, n. 1, p. 487–516, 2003. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.micro.57.030502.090709>>.

GROS, Francois et al. Unstable ribonucleic acid revealed by pulse labelling of Escherichia coli. *Nature*, v. 190, n. 4776, p. 581–585, 1961.

GUERRIER-TAKADA, Cecilia et al. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, v. 35, n. 3 PART 2, p. 849–857, 1983.

GUSAROV, Ivan; NUDLER, Evgeny. The mechanism of intrinsic transcription termination. *Molecular Cell*, v. 3, n. 4, p. 495–504, 1999.

- HA, Minju; KIM, V. Narry. Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*, v. 15, n. 8, p. 509–524, 2014. Disponível em: <<http://dx.doi.org/10.1038/nrm3838>>.
- HAASNOOT, Joost; WESTERHOUT, Ellen M.; BERKHOUT, Ben. RNA interference against viruses: Strike and counterstrike. *Nature Biotechnology*, v. 25, n. 12, p. 1435–1443, 2007.
- HAFEZ, Mohamed et al. A second eukaryotic group with mitochondrion-encoded tmRNA In silico identification and experimental confirmation. *RNA Biology*, v. 10, n. 7, p. 1117–1124, 2013.
- HARMS, Alexander et al. Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Molecular Cell*, p. 1–17, 2018. Disponível em: <<https://doi.org/10.1016/j.molcel.2018.01.003>>.
- HAYES, Finbarr; VAN MELDEREN, Laurence. Toxins-antitoxins: Diversity, evolution and function. *Critical Reviews in Biochemistry and Molecular Biology*, v. 46, n. 5, p. 386–408, 2011.
- HERNÁNDEZ, José A. et al. Identification of a furA cis antisense RNA in the cyanobacterium Anabaena sp. PCC 7120. *Journal of Molecular Biology*, v. 355, n. 3, p. 325–334, 2006.
- HEYER, Ruth et al. High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii*. *RNA Biology*, v. 9, n. 7, p. 1011–1018, 2012.
- HOAGLAND, M. B.; STEPHENSON M. L.; SCOTT, J. F.; HECHT L. I.; ZAMECNIK, P. C. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem*, v. 231, p. 241–257, 1958.
- HOBSON, David J. et al. RNA Polymerase II Collision Interrupts Convergent Transcription. *Molecular Cell*, v. 48, n. 3, p. 365–374, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.molcel.2012.08.027>>.
- HOCHSTRASSER, Megan L.; DOUDNA, Jennifer A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends in Biochemical Sciences*, v. 40, n. 1, p. 58–66, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.tibs.2014.10.007>>.
- HOLMQVIST, Erik; WAGNER, E Gerhart H. Impact of bacterial sRNAs in stress responses. *Biochemical Society transactions*, v. 0, n. September, p. BST20160363, 2017.
- HUBER, Florian et al. Protein Abundance Control by Non-coding Antisense Transcription. *Cell Reports*, v. 15, n. 12, p. 2625–2636, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.celrep.2016.05.043>>.
- INOUE, M. Antisense RNA: its functions and applications in gene regulation--a review. *Gene*, v. 72, p. 25–34, 1988.
- IRNOV, Irnov et al. Identification of regulatory RNAs in *Bacillus subtilis*. *Nucleic Acids Research*, v. 38, n. 19, p. 6637–6651, 2010.

- JÄGER, Dominik et al. Deep sequencing analysis of the *Methanosarcina mazei* G01 transcriptome in response to nitrogen availability. *PNAS*, 2009.
- JÄGER, Dominik et al. Primary transcriptome map of the hyperthermophilic archaeon *Thermococcus kodakarensis*. *BMC genomics*, v. 15, p. 684, 2014.
- JAIN, Samta; CAFORIO, Antonella; DRIESSEN, Arnold J.M. Biosynthesis of archaeal membrane ether lipids. *Frontiers in Microbiology*, v. 5, n. NOV, p. 1–16, 2014.
- JANSEN, Ruud. et al. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology*, v. 43, n. 6, p. 1565–1575, 2002. Disponível em: <<http://doi.wiley.com/10.1046/j.1365-2958.2002.02839.x>>.
- JENJAROENPUN, Piroon et al. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Research*, n. January, p. 1–15, 2018. Disponível em: <<http://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky014/4807344>>.
- JOHNSON, Jason M. et al. Dark matter in the genome: Evidence of widespread transcription detected by microarray tiling experiments. *Trends in Genetics*, v. 21, n. 2, p. 93–102, 2005.
- JØRGENSEN, Mikkel G. et al. HicA of *Escherichia coli* defines a novel family of translation-independent mRNA interferases in bacteria and archaea. *Journal of Bacteriology*, v. 191, n. 4, p. 1191–1199, 2009.
- KAMARTHAPU, Venu; NUDLER, Evgeny. Rethinking transcription coupled DNA repair. *Curr Opin Microbiol.*, v. 24, p. 15–20, 2016.
- KAPRANOV, Philipp et al. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, v. 316, n. 5830, p. 1484–1488, 2007.
- KAUR, Amardeep et al. A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome Research*, v. 16, n. 7, p. 841–854, 2006.
- KAWANO, Mitsuoki; ARAVIND, L.; STORZ, Gisela. An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. *Molecular Microbiology*, v. 64, n. 3, p. 738–754, 2007.
- KEREN, Iris et al. Specialized Persister Cells and the Mechanism of Multidrug Tolerance in *Escherichia coli*. *Journal of bacteriology*, v. 186, n. 24, p. 8172–8180, 2004.
- KIM, Daehwan; LANGMEAD, Ben; SALZBERG, Steven L. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, v. 12, n. 4, p. 357–360, 2015. Disponível em: <<http://dx.doi.org/10.1038/nmeth.3317>>.
- KIM, Donghyuk et al. Comparative Analysis of Regulatory Elements between *Escherichia coli* and *Klebsiella pneumoniae* by Genome-Wide Transcription Start Site Profiling. *PLoS Genetics*, v. 8, n. 8, p. e1002867, 2012. Disponível em: <<http://dx.plos.org/10.1371/journal.pgen.1002867>>.
- KLIEMT, Jana; SOPPA, Jorg. Chapter 10 Diverse Functions of Small RNAs (sRNAs) in Halophilic Archaea: From Non-coding Regulatory sRNAs to Microprotein-Encoding sRNAs.

RNA Metabolism and Gene Expression in Archaea. [S.l: s.n.], 2017. v. 32. p. 225–242. Disponível em: <<http://link.springer.com/10.1007/978-3-319-65795-0>>.

KOIDE, Tie et al. Prevalence of transcription promoters within archaeal operons and coding sequences. *Molecular systems biology*, NULL, v. 5, n. 285, p. 285, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19536208>>.

KOSKINENA, Kaisa et al. First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract ... crossm First Insights into the Diverse Human Archaeome: Specific Detection of Archaea in the Gastrointestinal Tract ,. v. 8, n. November, p. 1–17, 2017.

KROGER, C. et al. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proceedings of the National Academy of Sciences*, v. 109, n. 20, p. E1277–E1286, 2012. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.1201061109>>.

KRUGER, Kerstin; PFEIFER, Felicitas. Transcript Analysis of the c-vac Region and Differential Synthesis of the Two Regulatory Gas Vesicle Proteins GvpD and GvpE in *Halobacterium salinarum* PHH4. NULL, v. 178, n. 14, p. 4012–4019, 1996.

KUNG, J. T. Y.; COLOGNORI, D.; LEE, J. T. Long Noncoding RNAs: Past, Present, and Future. *Genetics*, v. 193, n. 3, p. 651–669, 2013. Disponível em: <<http://www.genetics.org/cgi/doi/10.1534/genetics.112.146704>>.

LARSEN, Helge; OMANG, Sissel; STEENSLAND, Heidi. On the gas vacuoles of the halobacteria. *Archiv für Mikrobiologie*, v. 59, n. 1–3, p. 197–203, 1967.

LASA, I et al. Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc.Natl.Acad.Sci.U.S.A*, v. 108, n. 1091–6490 (Electronic), p. 20172–20177, 2011. Disponível em: <c:%5CKARSTEN%5CPDFs%5CGrundlagen-PDFs%5CGrund-2011%5CLasa et al.-Genome-wide antisense transcription drives mRNA processing in bacteria.pdf>.

LASA, Inigo; TOLEDO-ARANA, Alejandro; GINGERAS, Thomas. An effort to make sense of antisense transcription in bacteria. *RNA Biol*, NULL, v. 9, n. 8, p. 1039–44, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3551857&tool=pmcentrez&rendertype=abstract>>.

LEIGH, John A. et al. Model organisms for genetics in the domain Archaea: Methanogens, halophiles, Thermococcales and Sulfolobales. *FEMS Microbiology Reviews*, v. 35, n. 4, p. 577–608, 2011.

LEVIN, Joshua Z. et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods*, v. 7, n. 9, p. 709–715, 2010. Disponível em: <<http://dx.doi.org/10.1038/nmeth.1491>>.

LI, Heng et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, v. 25, n. 16, p. 2078–2079, 2009.

LI, Jie et al. Global mapping transcriptional start sites revealed both transcriptional and post-transcriptional regulation of cold adaptation in the methanogenic archaeon *Methanobrevibacter*

psychrophilus. *Scientific reports*, v. 5, p. 9209, 2015. Disponível em: <<http://www.nature.com/srep/2015/150318/srep09209/full/srep09209.html>>.

LI, Zhongwei et al. RNA quality control: Degradation of defective transfer RNA. *EMBO Journal*, v. 21, n. 5, p. 1132–1138, 2002.

LISTER, Ryan et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, v. 462, n. 7271, p. 315–322, 2009. Disponível em: <<http://dx.doi.org/10.1038/nature08514>>.

LLORENS-RICO, Verónica et al. Bacterial antisense RNAs are mainly the product of transcriptional noise. *Science advances*, v. 2, n. 3, p. e1501363, 2016.

LORENZ, Ronny et al. ViennaRNA Package 2.0. *Algorithms for Molecular Biology*, v. 26, p. 1–14, 2011.

LOVE, Michael I.; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, v. 15, n. 12, p. 1–21, 2014.

LOWE, Todd M et al. Discovery of *Pyrobaculum* small RNA families with atypical pseudouridine guide RNA features. p. 402–411, 2012.

LYBECKER, Meghan et al. The double-stranded transcriptome of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, v. 111, n. 8, p. 3134–9, 2014. Disponível em: <<http://www.pnas.org/content/111/8/3134>>.

HARRINSON FC; KENNEDY, ME. The red discoloration of cured cod fish. *T Roy Soc Can Sect V* 16: 101–152, 1992.

MANDIN, Pierre; GUILLIER, Maude. Expanding control in bacteria: Interplay between small RNAs and transcriptional regulators to control gene expression. *Current Opinion in Microbiology*, v. 16, n. 2, p. 125–132, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.mib.2012.12.005>>.

MARCHLER-BAUER, Aron et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, v. 45, n. D1, p. D200–D203, 2017.

MÄRTENS, Birgit et al. Antisense regulation by transposon-derived RNAs in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *EMBO reports*, v. 14, n. 6, p. 527–33, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3674436&tool=pmcentrez&rendertype=abstract>>.

MARTIN, William F.; GARG, Sriram; ZIMORSKI, Verena. Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society B: Biological Sciences*, v. 370, n. 1678, 2015.

MATOS, Rute G. et al. The only exoribonuclease present in *Haloferax volcanii* has an unique response to temperature changes. *Biochimica et Biophysica Acta - General Subjects*, v. 1820, n. 10, p. 1543–1552, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.bbagen.2012.06.009>>.

MATOS, Rute Gonçalves; BARBAS, Ana; ARRAIANO, Cecília Maria. RNase R mutants elucidate the catalysis of structured RNA: RNA-binding domains select the RNAs targeted for

degradation. *The Biochemical journal*, v. 423, n. 2, p. 291–301, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19630750>>.

MATSUNAGA, F et al. In vivo interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin. *Proceedings of the National Academy of Sciences of the United States of America*, v. 98, p. 11152–11157, 2001.

MIKA, Franziska; HENGGE, Regine. Small regulatory RNAs in the control of motility and biofilm formation in *E. Coli* and *Salmonella*. *International Journal of Molecular Sciences*, v. 14, n. 3, p. 4560–4579, 2013.

MOISSEL-EICHINGER, Christine et al. Archaea Are Interactive Components of Complex Microbiomes. *Trends in Microbiology*, v. 26, n. 1, p. 70–85, 2018. Disponível em: <<http://dx.doi.org/10.1016/j.tim.2017.07.004>>.

NG, W V et al. Genome sequence of *Halobacterium* species NRC-1. *Proceedings of the National Academy of Sciences of the United States of America*, v. 97, n. 22, p. 12176–81, 2000. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17314&tool=pmcentrez&rendertype=abstract>>.

NICHOLSON, Allen W. Ribonuclease III mechanisms of double-stranded RNA cleavage. *Wiley Interdisciplinary Reviews: RNA*, v. 5, n. 1, p. 31–48, 2014.

NOLLER, H. Peptidyl Transferase: Protein, Ribonucleoprotein,. *JOURNAL OF BACTERIOLOGY*, v. 175, n. 17, p. 5297–5300, 1993.

NOLLER, H F; CHAIRES, J B. Functional modification of 16S ribosomal RNA by kethoxal. *Proceedings of the National Academy of Sciences of the United States of America*, v. 69, n. 11, p. 3115–3118, 1972.

OESTERHELT, Dieter; STOECKENIUS, Walther. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature New Biology*, v. 233, n. 39, p. 149–152, 1971.

OHTANI, Naoto et al. Cleavage of double-stranded RNA by RNase HI from a thermoacidophilic archaeon, *Sulfolobus tokodaii* 7. *Nucleic Acids Research*, v. 32, n. 19, p. 5809–5819, 2004a.

OHTANI, Naoto et al. Identification of the first archaeal Type 1 RNase H gene from *Halobacterium* sp. NRC-1: archaeal RNase HI can cleave an RNA–DNA junction. *Biochem. J*, v. 381, p. 795–802, 2004b.

OLIVAS, W M; MUHLRAD, D; PARKER, R. Analysis of the yeast genome: identification of new non-coding and small ORF-containing RNAs. *Nucleic acids research*, v. 25, n. 22, p. 4619–25, 1997. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=147069&tool=pmcentrez&rendertype=abstract>>.

ORELL, Alvaro et al. A regulatory RNA is involved in RNA duplex formation and biofilm regulation in *Sulfolobus acidocaldarius*. *Nucleic Acids Research*, v. 46, n. 9, p. 4794–4806, 2018. Disponível em: <<https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gky144/4909996>>.

OREN, A. Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *Journal of industrial microbiology & biotechnology*, v. 28, n. 1, p. 56–63, 2002.

OREN, A. Halophilic archaea on Earth and in space: growth and survival under extreme conditions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 372, n. 2030, p. 20140194–20140194, 2014. Disponível em: <<http://rsta.royalsocietypublishing.org/cgi/doi/10.1098/rsta.2014.0194>>.

OREN, Aharon. Taxonomy of the family Halobacteriaceae: A paradigm for changing concepts in prokaryote systematics. *International Journal of Systematic and Evolutionary Microbiology*, v. 62, n. 2, p. 263–271, 2012.

OREN, Aharon; TRÜPER, Hans. Anaerobic growth of halophilic archaeobacteria by reduction of dimethylsulfoxide and trimethylamine N-oxide. *FEMS Microbiology Letters*, v. 70, n. 1, p. 33–36, 1990.

OZSOLAK, Fatih et al. Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell*, v. 143, n. 6, p. 1018–1029, 2010. Disponível em: <<http://dx.doi.org/10.1016/j.cell.2010.11.020>>.

PAPENFORT, Kai; VOGEL, Jörg. Small RNA functions in carbon metabolism and virulence of enteric pathogens. *Frontiers in Cellular and Infection Microbiology*, v. 4, n. July, p. 1–12, 2014. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fcimb.2014.00091/abstract>>.

PECK, Ronald F.; DASSARMA, Shiladitya; KREBS, Mark P. Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Molecular Microbiology*, v. 35, n. 3, p. 667–676, 2000.

PELECHANO, Vicent; STEINMETZ, Lars M. Gene regulation by antisense transcription. *Nature Reviews Genetics*, v. 14, n. November, p. 880–93, 2013. Disponível em: <<http://www.nature.com/doi/10.1038/nrg3594>> <<http://www.ncbi.nlm.nih.gov/pubmed/24217315>>.

PERNITZSCH, Sandy R et al. An archaeal sRNA targeting cis - and trans -encoded mRNAs via two distinct domains. *Nature*, v. 40, n. 21, p. 10964–10979, 2012.

PETERS, Jason M.; VANGELOFF, Abbey D.; LANDICK, Robert. Bacterial transcription terminators: The RNA 3'-end chronicles. *Journal of Molecular Biology*, v. 412, n. 5, p. 793–813, 2011. Disponível em: <<http://dx.doi.org/10.1016/j.jmb.2011.03.036>>.

PFEIFER, Felicitas. Distribution, formation and regulation of gas vesicles. *Nature Reviews Microbiology*, v. 10, n. 10, p. 705–715, 2012. Disponível em: <<http://dx.doi.org/10.1038/nrmicro2834>>.

PFEIFER, Felicitas et al. Gas vesicle formation in halophilic archaea. *Archives of Microbiology*, v. 167, n. 5, p. 259–268, 1997.

PFEIFER, Felicitas. Haloarchaea and the Formation of Gas Vesicles. v. 4, p. 2015, 2015.

PFEIFER, Felicitas; GHAMRAN, Parvin. Plasmid pHH1 of *Halobacterium salinarum*: characterization of the replicon region, the gas vesicle gene cluster and insertion elements. *MGG Molecular & General Genetics*, v. 238, n. 1–2, p. 193–200, 1993.

- PFEIFFER, Friedhelm; OESTERHELT, Dieter. A Manual Curation Strategy to Improve Genome Annotation: Application to a Set of Haloarchael Genomes. *Life*, v. 5, n. 2, p. 1427–1444, 2015. Disponível em: <<http://www.mdpi.com/2075-1729/5/2/1427/>>.
- PORTNOY, Victoria; SCHUSTER, Gadi. RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R. *Nucleic Acids Research*, v. 34, n. 20, p. 5923–5931, 2006.
- PRASSE, Daniela; EHLERS, Claudia; SCHMITZ, Ruth A. Regulatory RNAs in archaea: first target identification in Methanoarchaea. v. 41, p. 344–349, 2013.
- QUINLAN, Aaron R. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*, v. 40, n. 6, p. 1301–1315, 2012.
- RACKER, Efraim; STOECKENIUS, Walther. Reconstitution of purple membrane vesicles catalyzing light-driven proton uptake and adenosine triphosphate formation. *Journal of Biological Chemistry*, v. 249, n. 2, p. 662–663, 1974.
- RAMÍREZ, Fidel et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic acids research*, v. 44, n. W1, p. W160–W165, 2016.
- RASMUSSEN, Simon; NIELSEN, Henrik Bjørn; JARMER, Hanne. The transcriptionally active regions in the genome of *Bacillus subtilis*. *Molecular Microbiology*, v. 73, n. 6, p. 1043–1057, 2009.
- RAY-SONI, Ananya; BELLECOURT, Michael J.; LANDICK, Robert. Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry*, v. 85, n. 1, p. 319–347, 2016. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev-biochem-060815-014844>>.
- REEVES, Matthew B. et al. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science*, v. 316, n. 5829, p. 1345–1348, 2007.
- REYNOLDS, Angela et al. Rational siRNA design for RNA interference. *Nature Biotechnology*, v. 22, n. 3, p. 326–330, 2004.
- RINGQUIST, Steven et al. Nature of the Ribosomal mRNA Track: Analysis of Ribosome-Binding Sites Containing Different Sequences and Secondary Structures. *Biochemistry*, v. 32, n. 38, p. 10254–10262, 1993.
- RINN, John L. lncRNAs : Linking RNA to Chromatin. *Cold Spring Harbor Perspectives in Biology*, p. 1–4, 2015.
- ROBERT, Xavier; GOUET, Patrice. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, v. 42, n. W1, p. 320–324, 2014.
- ROSENBLAD, Magnus Alm et al. Kinship in the SRP RNA family. *RNA biology*, v. 6, n. 5, p. 508–516, 2009.
- ROSS, J. A. et al. Hfq restructures RNA-IN and RNA-OUT and facilitates antisense pairing in the Tn10/IS10 system. *Rna*, v. 19, n. 5, p. 670–684, 2013. Disponível em: <<http://rnajournal.cshlp.org/cgi/doi/10.1261/rna.037747.112>>.

RUEPP, Andreas; SOPPA, Jörg. Fermentative arginine degradation in *Halobacterium salinarum* (formerly *Halobacterium halobium*): Genes, gene products, and transcripts of the *arcRACB* gene cluster. *Journal of Bacteriology*, v. 178, n. 16, p. 4942–4947, 1996.

SANTANGELO, Thomas J.; REEVE, John N. Archaeal RNA polymerase is sensitive to intrinsic termination directed by transcribed and remote sequences. *Journal of Molecular Biology*, v. 355, n. 2, p. 196–210, 2006.

SAPIENZA, Carmen; ROSE, Michael; DOOLITTLE, Ford. High-frequency genomic rearrangements involving archaeobacterial repeat sequence elements. *Nature*, v. 299, p. 182–185, 1982.

SAYED, Nour; JOUSSELIN, Ambre; FELDEN, Brice. A cis-antisense RNA acts in trans in *Staphylococcus aureus* to control translation of a human cytolytic peptide. *Nature Structural and Molecular Biology*, v. 19, n. 1, p. 105–113, 2012. Disponível em: <<http://dx.doi.org/10.1038/nsmb.2193>>.

SCHÄFER, Günter; PURSCHKE, Werner; SCHMIDT, Christian L. On the origin of respiration: Electron transport proteins from archaea to man. *FEMS Microbiology Reviews*, v. 18, n. 2–3, p. 173–188, 1996.

SCHIERLING, Karina et al. tRNA 3' end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*. *Journal of molecular biology*, v. 316, n. 4, p. 895–902, 2002.

SCHMID, Amy K et al. The anatomy of microbial cell state transitions in response to oxygen. The anatomy of microbial cell state transitions in response to oxygen. *Genome Research*, p. 1399–1413, 2007.

SEITZER, Phillip et al. A Monte Carlo-based framework enhances the discovery and interpretation of regulatory sequence motifs. *BMC Bioinformatics*, v. 13, n. 1, p. 1, 2012. Disponível em: <BMC Bioinformatics>.

SESTO, Nina et al. The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nature Reviews Microbiology*, v. 11, n. 2, p. 75–82, 2012. Disponível em: <<http://www.nature.com/doi/10.1038/nrmicro2934>>.

SHARMA, Cynthia M. et al. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, v. 464, n. 7286, p. 250–255, 2010.

SHARMA, Cynthia M.; VOGEL, Jörg. Differential RNA-seq: The approach behind and the biological insight gained. *Current Opinion in Microbiology*, v. 19, n. 1, p. 97–105, 2014.

SHEARWIN, Keith E; CALLEN, Benjamin P; EGAN, J Barry. Transcriptional interference – a crash course. *Trends in Genetics*, v. 21, n. 6, p. 339–345, 2010.

SHINE, J.; DALGARNO, L. The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences*, v. 71, n. 4, p. 1342–1346, 1974. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.71.4.1342>>.

SIEVERS, Fabian et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, v. 7, n. 539, 2011.

SILVA-ROCHA, Rafael et al. Development of New Modular Genetic Tools for Engineering the Halophilic Archaeon *Halobacterium salinarum*. *Plos One*, v. 10, n. 6, p. e0129215, 2015. Disponível em: <<http://dx.plos.org/10.1371/journal.pone.0129215>>.

SIMSEK, M et al. A transposable element from *Halobacterium halobium* which inactivates the bacteriorhodopsin gene. *Proc Natl Acad Sci U S A*, v. 79, n. 23, p. 7268–7272, 1982. Disponível em: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=6296826%5Chttp://www.pnas.org/content/79/23/7268.full.pdf>.

SLUPSKA, Malgorzata M. et al. Leaderless transcripts of the crenarchaeal hyperthermophile *Pyrobaculum aerophilum*. *Journal of Molecular Biology*, v. 309, n. 2, p. 347–360, 2001.

SMALDONE, Gregory T. et al. The FsrA sRNA and FbpB protein mediate the iron-dependent induction of the *Bacillus subtilis* lutABC iron-sulfur-containing oxidases. *Journal of Bacteriology*, v. 194, n. 10, p. 2586–2593, 2012.

SOPPA, Jörg et al. Small RNAs of the halophilic archaeon *Haloferax volcanii*: Figure 1. *Biochemical Society Transactions*, v. 37, n. 1, p. 133–136, 2009. Disponível em: <<http://biochemsoctrans.org/lookup/doi/10.1042/BST0370133>>.

ST.LAURENT, Georges; WAHLESTEDT, Claes; KAPRANOV, Philipp. The Landscape of long noncoding RNA classification. *Trends in Genetics*, v. 31, n. 5, p. 249–251, 2015. Disponível em: <<http://dx.doi.org/10.1016/j.tig.2015.03.007>>.

STAZIC, Damir; LINDELL, Debbie; STEGLICH, Claudia. Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Research*, v. 39, n. 11, p. 4890–4899, 2011.

STEENSELS, Jan; VERSTREPEN, Kevin J. Stop that Noise and Turn Up the Antisense Transcription. *Cell Reports*, NULL, v. 15, n. 12, p. 2575–2576, 2016. Disponível em: <<http://dx.doi.org/10.1016/j.celrep.2016.06.012>>.

STOLT, Pelle; ZILLIG, Wolfram. Structure specific ds/ss-RNase activity in the extreme halophile *Halobacterium salinarum*. *Nucleic Acids Research*, v. 21, n. 24, p. 5595–5599, 1993.

STOLT, Pelte; ZILLIG, Wolfram. Antisense RNA mediates transcriptional processing in an archaebacterium, indicating a novel kind of RNase activity. *Molecular Microbiology*, v. 7, n. 6, p. 875–882, 1993.

STORK, Michiel et al. Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *Journal of Bacteriology*, v. 189, n. 9, p. 3479–3488, 2007.

STRAUB, Julia et al. Small RNAs in Haloarchaea: Identification, differential expression and biological function. *RNA Biology*, v. 6, n. 3, p. 281–292, 2009. Disponível em: <<http://www.tandfonline.com/doi/abs/10.4161/rna.6.3.8357>>.

SUN, Yulong et al. Strategies to identify natural antisense transcripts. *Biochimie*, NULL, v. 132, p. 131–151, 2017. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S030090841630373X>>.

TADOKORO, Takashi; KANAYA, Shigenori. Ribonuclease H: Molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS Journal*, v. 276, n. 6, p. 1482–1493, 2009.

TANG, Thean Hock et al. Identification of novel non-coding RNAs as potential antisense regulators in the archaeon *Sulfolobus solfataricus*. *Molecular Microbiology*, NULL, v. 55, n. 2, p. 469–481, 2005.

TARASOV, Valery et al. A small basic protein from the *brz-brb* operon is involved in regulation of *bop* transcription in *Halobacterium salinarum*. *BMC Molecular Biology*, v. 12, n. 1, p. 42, 2011. Disponível em: <<http://www.biomedcentral.com/1471-2199/12/42>>.

THOMASON, Maureen K. et al. Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in *Escherichia coli*. *Journal of Bacteriology*, v. 197, n. 1, p. 18–28, 2015.

TIAN, Bin et al. The double-stranded-RNA-binding motif: interference and much more. *Nature reviews. Molecular cell biology*, v. 5, n. 12, p. 1013–1023, 2004.

TOLEDO-ARANA, Alejandro et al. The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*, v. 459, n. 7249, p. 950–956, 2009.

TURKARSLAN, Serdar et al. Niche adaptation by expansion and reprogramming of general transcription factors. *Molecular systems biology*, v. 7, n. 554, p. 554, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3261711&tool=pmcentrez&rendertype=abstract>>.

URBANOWSKI, Mark L.; STAUFFER, Lorraine T.; STAUFFER, George V. The *gcvB* gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in *Escherichia coli*. *Molecular Microbiology*, v. 37, n. 4, p. 856–868, 2000.

VAN, Phu T. et al. *Halobacterium salinarum* NRC-1 PeptideAtlas: strategies for targeted proteomics. *Proteome*, v. 7, n. 9, p. 3755–3764, 2009.

VAN BAKEL, Harm et al. Most “dark matter” transcripts are associated with known genes. *PLoS Biology*, NULL, v. 8, n. 5, 2010.

VANACORE, Daniela et al. Micrnas in prostate cancer: an overview. *Oncotarget*, v. 8, n. 30, p. 50240–50251, 2017.

VAN DER MEULEN, Sjoerd B.; DE JONG, Anne; KOK, Jan. Transcriptome landscape of *Lactococcus lactis* reveals many novel RNAs including a small regulatory RNA involved in carbon uptake and metabolism. *RNA Biology*, v. 13, n. 3, p. 353–366, 2016. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/15476286.2016.1146855>>.

VANHÉE-BROSSOLLET, Christine; VAQUERO, Catherine. Do natural antisense transcripts make sense in eukaryotes? *Gene*, v. 211, n. 1, p. 1–9, 1998.

VINCENT, Helen A.; DEUTSCHER, Murray P. The roles of individual domains of RNase R in substrate binding and exoribonuclease activity the nuclease domain is sufficient for digestion of structured RNA. *Journal of Biological Chemistry*, v. 284, n. 1, p. 486–494, 2009.

VOGEL, Christine; MARCOTTE, Edward M. Insights into regulation of protein abundance from proteomics and transcriptomics analyses. *Nature reviews genetics*, v. 13, n. 4, p. 227–232, 2013.

VOGEL, Jörg; LUISI, Ben F. Hfq and its constellation of RNA. *Nature Reviews Microbiology*, v. 9, n. 8, p. 578–589, 2011. Disponível em: <<http://www.nature.com/nrmicro/journal/v9/n8/full/nrmicro2615.html>> %5Cn<<http://www.nature.com/nrmicro/journal/v9/n8/pdf/nrmicro2615.pdf>>.

WADE, Joseph T.; GRAINGER, David C. Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nature Reviews Microbiology*, v. 12, n. 9, p. 647–653, 2014. Disponível em: <<http://dx.doi.org/10.1038/nrmicro3316>>.

WAGNER, E G H; SIMONS, R W. Antisense RNA Control in Bacteria, Phages, and Plasmids. *Annual Review of Microbiology*, v. 48, n. 1, p. 713–742, 1994. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.mi.48.100194.003433>>.

WAGNER, E Gerhart H; ROMBY, Pascale. *Small RNAs in Bacteria and Archaea: Who They Are, What They Do, and How They Do It*. [S.l.]: Elsevier Ltd, 2015. v. 90. Disponível em: <<http://dx.doi.org/10.1016/bs.adgen.2015.05.001>>.

WANG, Z. *et Al*. RNA-seq: A revolutionary tool for transcriptomics. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, v. 10, n. 1, p. 57–63, 2009.

WATANABE, Toshiaki *et al*. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature*, v. 453, n. 7194, p. 539–543, 2008.

WATERS, Lauren S; STORZ, Gisela. Regulatory RNAs in Bacteria. *Cell*, v. 136, n. 4, p. 615–628, 2009.

WEBB, Christopher J.; ZAKIAN, Virginia A. Telomerase RNA is more than a DNA template. *RNA Biology*, v. 13, n. 8, p. 683–689, 2016. Disponível em: <<http://dx.doi.org/10.1080/15476286.2016.1191725>>.

WELLNER, Karolin *et al*. Examining tRNA 3'-ends in Escherichia coli: A teamwork between CCA-adding enzyme, RNase T and RNase R. *RNA*, v. 49, n. 0, 2017.

WHITEHEAD, Kenia *et al*. An integrated systems approach for understanding cellular responses to gamma radiation. *Molecular Systems Biology*, v. 2, 2006.

WHITESIDE, Matthew D. *et al*. OrtholugeDB: A bacterial and archaeal orthology resource for improved comparative genomic analysis. *Nucleic Acids Research*, v. 41, n. D1, p. 366–376, 2013.

WIEDENHEFT, Blake; STERNBERG, Samuel H.; DOUDNA, Jennifer A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature*, v. 482, n. 7385, p. 331–338, 2012.

WILUSZ, Carol J; WILUSZ, Jeffrey. Eukaryotic Lsm proteins: lessons from bacteria. *Nature structural & molecular biology*, v. 12, n. 12, p. 1031–1036, 2005.

WOESE, C. R.; FOX, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, v. 74, n. 11, p. 5088–5090, 1977. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.74.11.5088>>.

- WOESE, C. R.; KANDLER, O.; WHEELIS, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, v. 87, n. 12, p. 4576–4579, 1990. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.87.12.4576>>.
- WURTMANN, Elisabeth J. et al. An evolutionarily conserved RNase-based mechanism for repression of transcriptional positive autoregulation. *Molecular Microbiology*, v. 92, n. 2, p. 369–382, 2014.
- WURTZEL, Omri et al. A single-base resolution map of an archaeal transcriptome. *Genome Research*, v. 20, n. 1, p. 133–141, 2010a.
- WURTZEL, Omri et al. A single-base resolution map of an archaeal transcriptome. *Genome Research*, NULL, v. 20, n. 1, p. 133–141, 2010b.
- YANOFSKY, C. Transcription attenuation. *Journal of Biological Chemistry*, v. 263, n. 2, p. 609–612, 1988.
- YU, Wenqiang et al. Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature*, v. 451, n. 7175, p. 202–206, 2008.
- ZARAMELA, Livia S. et al. Transcription start site associated RNAs (TSSaRNAs) are ubiquitous in all domains of life. *PLoS ONE*, v. 9, n. 9, 2014.
- ZHANG, Tao et al. Identification of a haloalkaliphilic and thermostable cellulase with improved ionic liquid tolerance. *Green Chemistry*, v. 13, n. 8, p. 2083–2090, 2011.
- ZHANG, Zhihong; DIETRICH, Fred S. Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5' SAGE. *Nucleic Acids Research*, v. 33, n. 9, p. 2838–2851, 2005.
- ZUO, Y; DEUTSCHER, M P. Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic acids research*, v. 29, n. 5, p. 1017–1026, 2001.
- ZUR BRUEGGE, Jennifer; EINSPANIER, Ralf; SHARBATI, Soroush. A Long Journey Ahead: Long Non-coding RNAs in Bacterial Infections. *Frontiers in Cellular and Infection Microbiology*, v. 7, n. March, p. 1–6, 2017. Disponível em: <<http://journal.frontiersin.org/article/10.3389/fcimb.2017.00095/full>>.

Apêndice

Apêndice A - Materiais Suplementares

1.1-Anotação e análises dos asRNAs

S1 – 2146 aTSSs mapeados

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S1.csv)

S2 – 1626 aTSSs filtrados por critérios de MFE

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S2.csv)

S3 – 846 asRNAs anotados

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S3.csv)

S4 – 145 asRNAs sobrepondo região 5' do gene sense

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S4.csv)

S5 – 118 asRNAs com expressao maior ou igual à do gene sense

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S5.csv)

S6 – 91 asRNAs interagindo co LSM

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S6.csv)

1.2- aTSSs em sistemas toxinas-antitoxinas do tipo II

S7 – *H. salinarum* NRC-1

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S7.csv)

S8 – *H. volcanii*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S8.csv)

S9 – *M. jannaschii*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S9.csv)

S10 – *T. kodakarensis*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S10.csv)

S11 – *E. coli*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S11.csv)

1.3-asRNAs diferencialmente expressos na curva de crescimento

S12 – 93 asRNAs diferencialmente expresso na transição de 17H-37H

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S12.csv)

S13 – 63 asRNAs diferencialmente expresso na transição de 37H-86H

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S13.csv)

S14 – Todos 144 asRNAs diferencialmente expressos em ao menos uma das transições

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S14.csv)

1.4-aTSSs conservados em *H. volcanii*

S15 – 160 asRNAs com aTSSs na mesma região em *H. volcanii*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S15.csv)

S16 – Genes ortólogos com mesmo tamanho e aTSSs na mesma região

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S16.csv)

1.5 - Genes diferencialmente expressos na linhagem *knockout* para o gene da RNase R

S17 – Genes upregulados na linhagem *knockout*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S17.csv)

S18 – Genes downregulados na linhagem *knockout*

(http://labisismi.fmrp.usp.br/Material_suplementar_dissertacao_JP/S18.csv)