**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

Junier Marrero Gutiérrez

Regulação pós-transcricional em *Halobacterium salinarum* NRC-1: caracterização do gene VNG_RS05855

Ribeirão Preto

2019

**UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**

**Regulação pós-transcricional em *Halobacterium salinarum* NRC-1: caracterização do
gene VNG_RS05855**

Tese apresentada à Faculdade de Medicina de Ribeirão Preto,
da Universidade de São Paulo, para obtenção do título de Dou-
tor em Ciências.

Área de Concentração: Bioinformática

Orientado: Junier Marrero Gutiérrez

Orientadora: Prof.^a Dr.^a Tie Koide

Co-orientador: Prof. Dr. Ricardo Z. N. Vêncio

Ribeirão Preto

2019

FICHA CATALOGRÁFICA

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.

Marrero-Gutiérrez, Junier.

Regulação pós-transcricional em *Halobacterium salinarum* NRC-1: caracterização do gene VNG_RS05855. / Junier Marrero Gutiérrez; orientadora Prof.^a Dr.^a Tie Koide e co-orientador Prof. Dr. Ricardo Zorzetto Nicoliello Vêncio, Ribeirão Preto, São Paulo, 2019. 216 folhas: Fig. 56, Tab.18.

Tese (Doutorado - Programa Interunidades de Pós-Graduação em Bioinformática: 1- Faculdade Medicina de Ribeirão Preto. 2- Instituto de Matemática e Estatística (IME)). Universidade de São Paulo, 2019

Área de Concentração: Bioinformática

1. metalo- β -lactamase. 2. RNases. 3. Ensaios fenotípicos. 4.RNA-seq. 5. Sítios de processamento do transcrito (TPS). 6. Estrutura secundária de RNA 7. Proteína Recombinante 8. Ensaios de atividade.

Este trabalho foi realizado com o suporte financeiro das seguintes entidades:

- Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP
- Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES- Demanda Social
- Fundação de Apoio ao Ensino, Pesquisa e Assistência do Hospital das Clínicas da FMRP– USP/FAEPA

FOLHA DE APROVAÇÃO

Nome: Marrero-Gutiérrez, Junier.

Título: Regulação pós-transcricional em *Halobacterium salinarum* NRC-1: caracterização do gene VNG_RS05855.

Tese apresentada à Faculdade de Medicina de Ribeirão Preto,
da Universidade de São Paulo, para obtenção do título de Dou-
tor em Ciências

Área de Concentração: Bioinformática

Orientadora: Prof.^a Dr.^a Tie Koide
Co-orientador: Prof. Dr. Ricardo Z. N. Vêncio

Aprovado em: ____ / ____ / ____

BANCA EXAMINADORA

Prof. Dr (a): _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr (a): _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr (a): _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr (a): _____ Instituição: _____

Julgamento: _____ Assinatura: _____

DEDICÁTORIA

Dedico

A minha mãe Martiza Gutiérrez Cáceres (1961-2017), Amo você!

A MORTE É A CURVA DA ESTRADA

A morte é a curva da estrada,

Morrer é só não ser visto.

Se escuto, eu te ouço a passada

Existir como eu existo.

A terra é feita de céu.

A mentira não tem ninho.

Nunca ninguém se perdeu.

Tudo é verdade e caminho.

Fernando Pessoa(Lisboa,1932)

AGRADECIMENTOS

Agradeço à Prof^a. Dr^a. Tie Koide pela oportunidade de integrar seu grupo de pesquisa, pelo incentivo, pelo apoio e por estar sempre disposta a discussões. Agradeço pelos conselhos e por me mostrar o caminho de como fazer e escrever ciência. Obrigada pela paciência, orientação, confiança e contribuição na minha formação profissional!

Agradeço ao Prof. Dr. Ricardo N. Z. Vêncio pela orientação, pelas constantes discussões e questionamentos, pelo apoio, pela confiança. Agradeço por me mostrar o útil de ser criterioso, focado, disciplinado e curioso na hora de fazer ciência, mas principalmente agradeço por ser amigo.

Agradeço ao Prof. Dr. Fernando Barroso, pela oportunidade de vir para o Brasil.

Agradeço a todo o corpo acadêmico e aos funcionários administrativos (antigos e atuais) do Programa de Interunidades de Pós-Graduação em Bioinformática, aos funcionários do Instituto de Matemática e Estatística da Universidade de São Paulo (IME), assim como aos funcionários do Departamento Bioquímica e Imunologia.

Agradeço à CAPES e à FAPESP pelos financiamentos da bolsa de estudo e do projeto.

Agradeço aos colaboradores do Institute for System Biology. Aos colaboradores do Laboratório Nacional de Luz Síncrotron (LNSL), em especial ao Prof. Dr. Douglas Galante. Aos colaboradores do Laboratório de Regulação da Expressão Gênica e Patogenicidade Bacteriana, em particular ao excelente Dr. Renato Rodrigues de Sousa Santos. Aos colaboradores do Laboratório de Genética Molecular e Bioinformática do HEMOCENTRO, em particular a outra excelente profissional Tecg.^o Kamila Peronni. Aos colaboradores do Laboratório Multiusuário de Microscopia Eletrônica: José A. Maulin, Ma. Tesera Picinto M. e Ma. Dolores Seabra F. Para todo este pessoal, as palavras de agradecimento, pela disposição, apoio e encorajamento, sempre seriam poucas, mas saibam que vou estando eternamente agradecido.

Agradeço a Sílvia Epifânio por todo apoio, pela colaboração, por todos os meios de cultura, géis, soluções, curvas de crescimento. Agradeço imensamente a atenção, o carinho, preocupação, e os tanto conselhos de tantas conversas. Muito obrigado!

Agradeço aos amigos e colegas do Laboratório de Biologia Sistêmica de Microrganismos (Labisismi) e do Laboratório de Processamento de Informação Biológica (LabPIB), pelos

que desenvolvi um imensurável respeito e admiração. Todos fizeram parte desta etapa de amadurecimento pessoal e profissional. Aos que estão aqui, e aos que não, eu agradeço.

Agradeço aos membros da banca por ter aceito apoiar o desenvolvimento deste trabalho, pelas críticas e sugestões feitas a ele e a mim, pelo bom censo, pela confiança, pelo fato de ter que ler estas páginas e ainda mostrar entusiasmo. Agradeço por favorecer meu crescimento como pesquisador.

Agradeço a meus amigos cubanos, esses que estão pelo mundo todo, aos amigos brasileiros e africanos: José Enrique Eirez, Malena Martinez, Alieski Cruz Ramirez, Daniel Ferrer Vinlas, Fabiana Fuentes León, Jeffrey Álvarez Masson, Carolina Teixeira, Amr Galal e Yagoub Ali Abraham. Em especial a meu eterno professor o Dr. Angel Sánchez Lamar. Em meu peito, bem no meio, vocês tem um lugar.

Agradeço a meus pais, Javier Marrero Fraga e Martiza Gutiérrez Cáceres, pelas lições de humildade, honestidade, por terem me ensinado o que há de mais importante na vida. Agradeço o seu apoio incondicional, e o seu amor. Em especial a você, minha mãe, que ouço, sinto, amo e penso sempre. Sabes, “a morte é mentira quando se tem cumprido bem a obra da vida” (José Martí, herói cubano), por isso estás comigo, por isso estou aqui.

Agradeço a minha esposa, Cibele Cirino de Souza, seu apoio, seu incentivo. Agradeço pelos sonhos compartilhados, por estar a meu lado em todos os momentos. Agradeço por me enxergar, por falar “sim”. Você é o que faltava em mim. Amo você!

Agradeço a minha família: minha irmã, Maviela Marrero, e minha sobrinha, Laritza, a minha avô, Felicita, que insiste me olhar como se ainda fosse uma criança. Agradeço a meus irmãos, Javier e Olyedis, e a seus filhos também. Eu sou fruto de todos vocês.

Agradeço a família de minha esposa, que também é minha família. Tarciso Gaspar e Maria Cirino, que acolheram este imigrante como um filho a mais. Agradeço a meus cunhados Tarcísio G.S. Filho; Sílvia C.S. Schiavo e seu esposo Haroldo Schiavo. Todos exemplos de dedicação e humildade. Obrigado.

Agradeço a Cuba, minha Pátria, e ao Brasil, minha casa.

RESUMO

JUNIER, M. G. **Regulação pós-transcricional em *Halobacterium salinarum* NRC-1: caracterização do gene VNG_RS05855**. 2019. 225f. Teses (Doutorado) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

A família das enzimas metalo- β -lactamases compreende um grupo de proteínas com participação no metabolismo dos ácidos nucleicos. Nos genomas procariotos, muitas proteínas estão anotadas como potenciais metalo- β -lactamases, porém sem função comprovada. Este é o caso do gene VNG_RS05855 de *Halobacterium salinarum* NRC-1, organismo em que os principais atores da regulação pós-transcricional, as RNases, também são pouco conhecidos. Nesta pesquisa, relatamos a primeira caracterização da metalo- β -lactamase VNG_RS05855 da arqueia halofílica *H. salinarum* NRC-1 no contexto da regulação pós-transcricional, mediante a integração de abordagens experimentais e computacionais. VNG_RS05855 compartilha similaridade de sequência com proteínas descritas com atividade ribonucleolítica, porém a ausência de domínios específicos (exosite) coloca em dúvida a afirmação desta função. O gene apresenta uma expressão relativa basal durante o crescimento, que se altera diante de perturbações ambientais e genéticas, e é correlacionada a outras RNases como ocorre com RNase de outros organismos. Embora a ausência do gene VNG_RS05855 não tenha afetado nem a viabilidade nem o crescimento da linhagem nocaute (Δ VNG_RS05855), foram observadas ligeiras diferenças fenotípicas em relação a linhagem controle frente a baixas concentrações de metais de transição (Zn^{2+} , Cu^{2+} , Fe^{2+} , Mn^{2+} , Co^{2+}), ao metalóide NaAsO_2 , ao agente oxidante H_2O_2 e a variações de sais do meio de cultura. Outros fatores que poderiam ser responsáveis por essas diferenças como a liberação de vesículas de gás e a produção de bacteriorodopsina foram descartados. Também foram identificadas diferenças no pH diante das transições metabólicas promovidas por mudanças na disponibilidade de nutrientes e luz. Ao analisar a sobrevivência da linhagem nocaute em concentrações semi-letais de Cu^{2+} , AsO_2^- e PQT não observamos diferenças em relação a linhagem controle, sugerindo que no prévio condicionamento ao estresse, VNG_RS05855 não participe nas respostas adaptativas. De acordo com análises do transcritoma, a ausência do gene VNG_RS05855 afetou a expressão de 117 genes, 44% deles sem função conhecida. Entre os que se conhece a função se destacam os transcritos relacionados às vesículas de gás e transporte. Na produção das vesículas de gás, as análises da cobertura de reads sugerem que a participação de VNG_RS05855 esteja relacionada a regulação antisense. A expressão de genes relacionados ao transporte nos indica que o metabolismo do RNA é complexo e o impacto do VNG_RS05855 na regulação pós-transcricional, heterogêneo. Dentre os transportadores com expressão afetada estão os relacionados às respostas a arsênico, sugerindo o seu papel na resposta adaptativa frente a este metalóide. A análise de sítios de processamento utilizando dados de transcritoma (TPS-transcriptional processing sites) indica a influência diferencial no processamento dos transcritos dos replicons de *H. salinarum* NRC-1, com VNG_RS05855 atuando menos nos transcritos derivados do cromossomo que nos dos megaplasmídeos. Além disso, observamos o processamento preferencial de transcritos longos, apoiando a expressão diferencial de transcritos dentro de unidades policistrônicas. A estruturação dos RNAs poderia contribuir para sinalizar o grau de processamento promovido por VNG_RS05855. A obtenção da proteína heteróloga e os testes de atividade mostraram com clareza o papel ribonucleolítico in vitro do VNG_RS05855, tornando-se o primeiro retrato deste tipo de atividade em proteínas metalo- β -lactamases que carecem de domínio exosite na classe Halobacteria.

Palavras-chaves: *Halobacterium salinarum* NRC-1, metalo- β -lactamases, VNG_RS05855, RNases, regulação pós-transcricional, asRNA, ensaios fenotípicos, RNA-seq, bioinformática, atividade in vitro.

ABSTRACT

JUNIER, M. G. **Post-transcriptional regulation in *Halobacterium salinarum* NRC-1: characterization of VNG_RS05855 gene.** 2019,225p. Theses (Doctorate) - Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto.

The family of metallo- β -lactamase enzymes comprises a group of diverse proteins, some of them with participation in nucleic acid metabolism. In prokaryotic genomes many proteins are annotated as potential metallo- β -lactamases, but without proven function. This is the case of gene VNG_RS05855 from *Halobacterium salinarum* NRC-1, an organism in which the main actors of post-transcriptional regulation, the RNases, are still poorly characterized. In this research, we report the first characterization of the metallo- β -lactamase VNG_RS05855 in the context of post-transcriptional regulation of *H. salinarum* NRC-1 halophilic archaeon by integrating experimental and computational approaches. VNG_RS05855 has sequence similarity to proteins described with ribonuclease activity, but the absence of specific domains (exosite) casts doubt of this possible function. VNG_RS05855 presents constitutive expression during growth and environmental and genetic perturbations its expression levels correlates with other RNases, as observed in other organisms. Although the absence of the VNG_RS05855 gene did not affect either the viability and growth of the knockout strain, we observed slight phenotypic differences under low concentrations of transition metals (Zn^{2+} , Cu^{2+} , Fe^{2+} , Mn^{2+} , Co^{2+}), a metalloid (NaAsO_2), an oxidizing agent (H_2O_2) and salt variations of the culture medium. Differences were also observed in the pH of the growth media before the metabolic transitions promoted by changes in nutrient availability and light. The results suggest the participation of VNG_RS05855 in adaptive responses to small environmental changes. The phenotypes were observed in the final stages of growth where other factors could also be related to the observed differences, such as gas vesicle release and bacteriorhodopsin production. However, our experiments showed that these factors do not affect the observed phenotype. Nevertheless, when analyzing the survival of the knockout strain in semi-lethal concentrations of Cu^{2+} , AsO_2^- and PQT, we noticed that there were no differences regarding the control strain, suggesting that in the previous stress conditioning, VNG_RS05855 does not participate in the adaptive responses. According to transcriptome analysis, the absence of the VNG_RS05855 affected the expression of 117 genes, 44% of them with unknown function. Among the genes with assigned function, we found transcripts related to gas vesicles and transport. In the production of gas vesicles, analysis of reads coverage suggests that the participation of VNG_RS05855 is related to antisense regulation. The expression of transport-related genes lead us to suggest that RNA metabolism is complex and the impact of VNG_RS05855 on post-transcriptional regulation is heterogeneous. Among the differentially expressed transporters we found genes related to arsenic response, suggesting the role of VNG_RS05855 in the adaptive response to this metalloid. Analysis of transcriptional processing sites (TPS) showed differential influence on the processing of *H. salinarum* NRC-1 replicons, a smaller density of TPS was found on chromosome-derived transcripts than on megaplasmids. In addition, we observed the preferential processing of long transcripts, supporting the differential expression of transcripts within polycistronic units. RNA secondary structure could contribute to indicate the degree of processing promoted by the activity of VNG_RS05855. We were able to express the heterologous protein of this gene and test its activity. The results clearly show its in vitro ribonucleolytic role, showing for the first time of this kind of activity in metallo- β -lactamase protein lacking exosite domain in Halobacteria class.

Keywords: *Halobacterium salinarum* NRC-1, metallo- β -lactamases, VNG_RS05855, RNases, post-transcriptional regulation, asRNA, phenotypic assays, RNA-seq, bioinformatics, in vitro

assays.

Lista de Figuras

1.1	Representação esquemática de proteínas com domínio metalo- β -lactamase relacionadas ao metabolismo dos ácidos nucleicos.	6
1.2	Os três domínios da vida	9
1.3	<i>Halobacterium salinarum</i> NRC-1	13
1.4	Acoplamento quimiosmótico entre a energia solar, a bacteriorodopsina na produção de ATP	14
3.1	Esquema representativo da técnica de dRNA-seq (differential RNA-seq)	28
3.2	Representação esquemática na abordagem <i>in house</i> para identificar sítios de processamento nos transcritos.	35
4.1	Alinhamento múltiplo da sequência de aminoácidos do gene VNG_RS05855 .	43
4.2	Perfis de expressão relativa de genes com possível atividade ribonucleolítica em <i>H. salinarum</i> NRC-1	46
4.3	Correlação da expressão dos genes anotados como RNases em <i>H. salinarum</i> NRC-1	49
4.4	Média das variações da absorbância a 600nm (OD) e da densidade celular das diferentes linhagens	52
4.5	Curvas de crescimento sob a influência de metais de transição	54
4.6	Curvas de crescimento sob agentes oxidantes	55
4.7	Curvas de crescimento sob metaloide	56
4.8	Curvas de crescimento sob mudanças de sal do meio complexo (CM)	57
4.9	Conjunto total de medições de absorbância a 600 nm das linhagens nas condições de referência del crescimento	59
4.10	Conjunto total de medições de absorbância a 600 nm das linhagens sob os diferentes estresses	61
4.11	Medição da absorbância a diferentes comprimentos de onda.	63
4.12	Medições de pH das linhagens	65
4.13	Medições de pH das linhagens diluídas em solução basal	66
4.14	Medições de pH das linhagens diluídas em meio CM	66
4.15	Espectroscopia Raman das linhagens após de excitação das amostras a 532 nm. .	69
4.16	Espectroscopia Raman das linhagens após excitação das amostras a 785 nm . .	70
4.17	Intensidades Raman dos biomarcadores de <i>H. salinarum</i> NRC-1 das linhagens às 41 e 95 horas de crescimento.	72
4.18	Microfotografias das linhagens às 17, 41, 65 e 95 horas do crescimento	74
4.19	Ensaio de sobrevivência das linhagens submetido a condição semi-letal de CuSO_4	75
4.20	Ensaio de sobrevivência das linhagens submetido a condição semi-letal de NaAsO_2	76
4.21	Ensaio de sobrevivência das linhagens submetido a condição semi-letal de PQT.	76
4.22	Cobertura de <i>reads</i> do operon das vesículas de gás	90
4.23	Cobertura de <i>reads</i> do operon de arsênico	91
4.24	Cobertura de <i>reads</i> das operons da família de transportadores ABC	92
4.25	Cobertura de <i>reads</i> de reguladores transcricionais	93

4.26	Sítios de processamento identificadas com metodologia desenvolvida “in-house”	94
4.27	Sítios de processamento dos transcritos (TPS) identificados pelo metodo “in-house” associados a atividade do VNG_RS05855.	95
4.28	Localização dos TPS em transcritos de rRNA e tRNA	96
4.29	Gráficos de MA-plot para amostras preparadas com os protocolo A) TruSeq-Stranded e B) TruSeq-Small	99
4.30	Tamanho dos genes com expressão diferencial identificados a partir de preparos diferentes de bibliotecas.	100
4.31	Análises dos resultados do dRNA-seq com a ferramenta TSSAR stand-alone (Amman et al., 2014) para determinar inícios da transcrição (TSS)	109
4.32	Análises dos resultados do dRNA-seq com a ferramenta TSSAR stand-alone (Amman et al., 2014) para determinar sítios de processamento dos transcritos (TPS) nas linhagens	110
4.33	Sítios de processamento pós-transcricional (TPS) do produto proteico do gene VNG_RS05855.	112
4.34	Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 no transportadores da família de ABC	114
4.35	Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 no operon das vesículas de gás.	115
4.36	Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 em reguladores transcricionais e na halorodopsina.	116
4.37	Número de genes com expressão diferencial (cor: magenta) que tem associado TSP do VNG_RS05855	118
4.38	Histograma das mínimas energias livres (mfe-kcal/mol-) para o conjunto de sequências que contêm os TPS e a curva da distribuição normal dessa variável.	119
4.39	Análises das estruturas secundárias das sequências que contem TPS primários	122
4.40	Análises das estruturas secundárias das sequências contêm os TPS internos.	123
4.41	Análises das estruturas secundárias das sequências que contêm os TPS anti-sense.	124
4.42	Gel de poliacrilamida 15% com as amostras do processo de purificação da proteína recombinante VNG_RS05855.	125
4.43	Eletoferograma dos RNA total das linhagens controle e nocaute	126
4.44	Eletoferograma dos RNA pequeno das linhagens controle e nocaute	127
4.45	Eletoferograma dos RNA total das linhagens controle e nocaute com kit de RNA-small	128
4.46	Atividade da proteína recombinante frente ao RNA total	130
4.47	Atividade da proteína recombinante frente ao RNA pequeno	130
4.48	Atividade da proteína recombinante frente ao RNA total com RNA-small	131
4.49	RNA total das linhagens após de incubar com tampão de reação	132
4.50	RNA total e pequeno frente a RNases comerciais	133

Lista de Tabelas

1.1	Principais famílias de enzimas envolvidas na degradação e processamento de RNAs celulares (adaptado de (Arraiano et al., 2013))	4
1.2	Características apresentadas pelas Bactérias, Arqueias e Eucariotos	8
3.1	Experimentos de <i>microarray</i> e <i>tiling array</i> de <i>H. salinarum</i> NRC-1 utilizados neste trabalho	19
3.2	Linhagens de <i>H. salinarum</i> utilizadas nesta pesquisa	20
3.3	<i>Primers</i> utilizados neste trabalho	27
3.4	Resumo do estudo das técnicas de RNA-seq para estudar o transcrito do Δ VNG_RS05855	31
4.1	Organismos que apresentam genes com maior percentual de identidade de sequências com o gene VNG_RS05855.	41
4.2	Dez proteínas com os menores <i>E-value</i> que apresentam os domínios conservados identificados no gene VNG_RS05855.	41
4.3	Números de reads sequenciados e alinhados nas distintas amostras das linhagens	78
4.4	Genes com expressão diferencial na linhagem nocaute VNG_RS05855.	80
4.5	(continuação): Genes com expressão diferencial na linhagem nocaute VNG_RS05855	81
4.6	Termos <i>GO</i> enriquecidos nos genes com expressão diferencial (DESeq2) na linhagem nocaute VNG_RS05855	82
4.7	Termos <i>GO</i> enriquecidos nos genes com expressão diferencial estimados por análises MA-plot a partir do protocolo TruSeq-Stranded.	101
4.8	Termos <i>GO</i> enriquecidos nos genes com expressão diferencial estimados por análises MA-plot a partir do protocolo TruSeq-Small.	102
4.9	Genes com expressão diferencial, pela análise do DESeq2 membros de operons.	104
4.10	Seleção de 43 genes de um total de 139 com expressão diferencial (Anexo 7.2, restantes genes), segundo as análises utilizando MA-plot membros de operons.	105
4.11	Genes com expressão diferencial, segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Small membros de operons	106
4.12	Termos <i>GO</i> enriquecidos nos genes com TPS do produto proteico do gene VNG_RS05855.	118

Lista de Abreviações

Transportadores ABC - ATP-binding cassette, transportadores dependentes de ATP,

asRNA - antisense RNA; RNA antisenso

aTSS - antisense transcript start site; início de início da transcrição antisenso

cDNA - complementary DNA – DNA complementar

CM - complex media; meio complexo

CoSO₄ . 7H₂O - Sulfato de cobalto hepta hidratado

CuSO₄ . 7H₂O -Sulfato de cobre hepta hidratado

dRNA-seq - differential RNA-seq; RNA-seq diferencial; sequenciamento de RNA primário

FC - representa o valor de expressão de cada gene na amostra teste normalizado pelo valor da amostra controle, fold change

FeSO₄ . 7H₂O - Sulfato de ferro hepta hidratado

fwd - forward

GEO - Gene Expression Omnibus

GFF - general feature file

GGB - Gaggle genome browser

GO - Gene ontology

H₂O₂ - peróxido de hidrogênio

LB - Luria bertani

MBL - metalo-β-lactamase

MFE - minimum free energy; mínima energia livre

MgSO₄ . 7H₂O - Sulfato de magnésio hepta hidratado

MnSO₄ . 7H₂O - Sulfato de manganês hepta hidratado

mRNA - messenger RNA; RNA mensajero

NaAsO₂ - arsenito de sódio

NCBI - National Center for Biotechnology Information

ncRNA - non coding RNA; RNA não-codificante

OD - optical density; densidade óptica

ORF - Open Reading Frame; janela aberta de leitura

PAGE – gel de poliacrilamida

PE - paired-end

PQT - Paraquat; 1,1'-dimetil-4,4'-bipiridina-dicloreto

rev - reverse

RNA-seq - RNA sequencing; sequenciamento de RNA

rRNA - ribosomal RNA; RNA ribossomal

SDS – dodecil-sulfato de sódio

TEX - Terminator 5'-Phosphate-Dependent Exonuclease

tRNA – transfer RNA; RNA transportador

TSS - transcript start site; sítios de início de transcrição

TPS - transcript processing site; sítio de processamento do transcrito

UTR - untranslated region; região não traduzida

ZnSO₄ . 7H₂O – Sulfato de zinco hepta hidratado

Sumário

1	Introdução	1
1.1	Processamento de RNAs	1
1.2	RNases da superfamília das metalo- β -lactamase	5
1.3	Archaea: o terceiro domínio da vida	7
1.4	Regulação transcricional e pós-transcricional em Archaeas	9
1.5	<i>Halobacterium salinarum</i> NRC-1	12
1.6	Regulação pós-transcricional em arqueias halofílicas	15
2	Objetivos	18
2.1	Geral	18
2.2	Específicos	18
3	Materiais e Métodos	19
3.1	Banco de dados	19
3.2	Dados públicos de microarray e tiling array	19
3.3	Linhas e condições de cultivo	20
3.4	Ensaio fenotípicos	21
3.4.1	Curvas de crescimento	21
3.4.2	Determinação da densidade celular	22
3.4.3	Medições da absorvância em diferentes comprimentos de onda	22
3.4.4	Medições de pH	23
3.4.5	Espectroscopia RAMAN	23
3.4.6	Microscopia eletrônica de transmissão	24
3.4.7	Ensaio de sobrevivência	25
3.5	Estudo do transcrito	25
3.5.1	Extração do RNA total e pequenos RNAs	25
3.5.2	Tratamento com DNases	26
3.5.3	Enriquecimento dos RNAs primários	27
3.5.3.1	Software para a identificação em larga escala de inícios da transcrição	27
3.5.4	Depleção de RNA ribossomal	29
3.5.5	Fragmentação do RNA depletado	29
3.5.6	Preparação das bibliotecas de RNA-seq	30
3.5.6.1	TruSeq Small	30
3.5.6.2	TruSeq mRNA Stranded	31
3.5.7	Quantificação das bibliotecas	32
3.5.8	Sequenciamento das bibliotecas	32
3.5.9	Processamento dos dados de sequenciamento	33
3.6	Exploração de cobertura de reads obtida do sequenciamento de amostras de RNA total depletado de rRNA sem fragmentar	34
3.7	Obtenção da proteína recombinante do gene VNG_RS05855 e ensaios de atividade	35
3.7.1	Expressão e purificação da proteína recombinante	35
3.7.2	Ensaio de atividade da proteína recombinante	38

4	Resultados	40
4.1	A sequência proteica do gene VNG_RS05855 conserva os motivos das proteínas metalo- β -lactamase com função ribonucleolítica	40
4.2	Os níveis de expressão relativa das RNases de <i>H. salinarum</i> NRC-1 são baixos, segundo as análises nos dados públicos de microarray e tiling array	44
4.3	Ensaio fenotípicos	50
4.3.1	A linhagem nocaute do gene VNG_RS05855 é viável e apresenta crescimento similar à linhagem controle	51
4.3.2	Sob estresses o crescimento da linhagem nocaute apresenta sutis diferenças em relação à linhagem controle	52
4.3.2.1	Metais de transição	53
4.3.2.2	Estresse oxidativo	53
4.3.2.3	Arsênico	55
4.3.2.4	Concentração de sal	56
4.3.3	As diferenças nas medições de absorvância se concentram nas fases finais do crescimento	57
4.3.4	As diferenças de absorvância nas etapas finais do crescimento não está associada com o acúmulo diferencial de bacteriorodopsina entre as linhagens	62
4.3.5	Ausência do gene VNG_RS05855 influencia as transições metabólicas diante de mudanças na disponibilidade de nutrientes	63
4.3.6	Os principais biomarcadores <i>H. salinarum</i> NRC-1 não são afetados pela ausência do gene VNG_RS05855	66
4.3.7	A produção de vesículas de gás entre as linhagens nocaute e controle não apresentam diferenças relevantes	73
4.3.8	Diante das concentrações semi-letais de agentes estressantes, as linhagens têm o mesmo nível de sobrevivência	74
4.4	Transcritoma das linhagens controle e nocaute	77
4.4.1	Análise dos dados de RNAseq sem fragmentação	78
4.4.1.1	No transcritoma da linhagem nocaute a expressão de 117 genes é alterada	78
4.4.1.2	Processos biológicos relacionados aos genes com expressão diferencial	82
4.4.1.2.1	Organização de vesículas de gás	83
4.4.1.2.2	Respostas a arsênico	84
4.4.1.2.3	Respostas a estímulos	85
4.4.1.2.3.1	Transportadores	85
4.4.1.2.3.2	Reguladores transcricionais	87
4.4.1.2.3.3	Outros genes	88
4.4.1.3	Estruturas dos sinais de expressão gênica	89
4.4.1.3.1	Perfil dos sinais de expressão dos genes associados a processos biológicos	89
4.4.1.3.2	Rastros do processamento desigual entre as linhagens por diferenças nos sinais de expressão	93
4.4.2	Análise dos dados de RNAseq com fragmentação	96
4.4.2.1	Papel de VNG_RS05855 no processamento de transcritos longos relacionados a diversos processos biológicos	97

4.4.2.2	Estabilidade de unidades policistrônicas podem relacionar-se com a ausência do gene VNG_RS05855	102
4.5	Análises em larga escala dos sítios que marcam os inícios de transcrição e de processamentos do RNA (dRNAseq)	106
4.5.1	Identificação dos sítios de início da transcrição (TSS)	107
4.5.2	Identificação dos sítios processamento do transcrito (TPS)	109
4.5.3	Possíveis TPS do produto proteico do gene VNG_RS05855	111
4.5.3.1	TPS nos genes com expressão diferencial	113
4.5.3.2	Processos biológicos associados aos genes com TPS	117
4.5.3.3	Estrutura secundária das sequências que contem os TPS	119
4.5.3.3.1	Não há conservação nas estruturas das sequências que contem os TPS associados ao gene VNG_RS05855	120
4.6	Expressão da proteína recombinante do gene VNG_RS05855 em <i>E. coli</i>	124
4.6.1	Os conjuntos dos RNA das linhagens não são claramente diferentes	125
4.6.2	A proteína recombinante processa substratos de RNA in vitro	129
4.6.3	Perfil do processamento da proteína recombinante é similar ao das RNases R e A	132
5	Discussão	134
5.1	Características da sequência VNG_RS05855 sugerem que seu produto proteico seja uma RNase	134
5.2	A ausência do gene VNG_RS05855 não afeta o crescimento da linhagem nocaute e influencia de forma sutil a resposta perante estresses	135
5.3	Na linhagem nocaute a produção de bacteriorodopsina não é afetada, porém diante de mudanças na disponibilidade de nutrientes, as transições metabólicas são perturbadas	138
5.4	A ausência gene VNG_RS05855 não perturba a produção de carotenóides nem vesículas de gás	141
5.5	VNG_RS05855 não desempenha um papel-chave na sobrevivência de <i>H. salinarum</i> NRC-1 frente a Cu^{2+} , AsO_2^- e PQT	142
5.6	A regulação por RNA antisense de genes relacionados ao operon das vesículas de gás pode ser afetada na ausência do gene VNG_RS05855	143
5.7	VNG_RS05855 pode estar envolvido na regulação da expressão de genes relacionados à família de transportadores ABC	145
5.8	A localização do produto proteico do gene VNG_RS05855 pode ser restrita a determinados espaços celulares	147
5.9	O gene VNG_RS05855 influencia de maneira diferencial o processamento dos transcritos dos distintos replicons de <i>H. salinarum</i> NRC-1	148
5.10	VNG_RS05855 está envolvido no processamento de transcritos longos	150
5.11	Unidades policistrônicas de <i>H. salinarum</i> NRC-1 são diferencialmente processadas na linhagem nocaute do gene VNG_RS05855	150
5.12	A estruturação do RNA dita o grau de processamento nos transcritos de <i>H. salinarum</i> NRC-1	152
5.13	Processos biológicos afetados na ausência do gene VNG_RS05855	154
5.14	Atividade ribonuclease da proteína codificada no gene VNG_RS05855	155

6	Conclusões	157
7	Anexos	161
7.1	Curvas de crescimento das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) submetidas a diferentes estresses	161
7.2	Genes com expressão diferencial, segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Stranded, identificados como possíveis membros de operons	168
7.3	OS DADOS DEPOSITADOS EM:	170

1 Introdução

1.1 Processamento de RNAs

A expressão de informações genéticas depende de três tipos principais de RNAs: RNAs mensageiros (mRNAs), intermediários na definição do proteoma; RNAs ribossomais (rRNA) e RNAs transportadores (tRNAs) que são as moléculas efetoras que atuam juntas para decodificar a informação contida na sequência do mRNA; e finalmente o conjunto abundante de outros RNAs não codificantes (ncRNAs) que podem apresentar papel-chave na regulação pós-transcricional. A função destes ncRNA abrange desde regular a meias-vidas de alguns mRNAs (sRNA/microRNA/siRNA) (Carthew, Richard W. and Sontheimer, 2009), favorecer processos de splicing (snRNAs) e de edição dos mRNA (snoRNA -box C/D e H/ACA-) (Xue Hai et al., 2005), assim com regular a taxa de tradução (Brantl, 2002; Sesto et al., 2013).

Os níveis de expressão de um RNA são determinados pelas taxas de transcrição e degradação. Além disso, RNA recém formados, ou primários, sofrem múltiplas modificações para alcançar seu estado funcional. Esta diversidade de eventos ocorre pela ação coordenada de um conjunto de enzimas de síntese, degradação e maturação, que são as encarregadas de monitorar com precisão os níveis e a qualidade do RNA (Arraiano et al., 2013). Um grupo de enzimas, diretamente responsáveis em alguns destes processos, são as ribonucleases (RNases) que catalisam a ruptura exo- ou endo-nucleolítica das ligações fosfodiéster do RNA. A maioria das RNases conhecidas são enzimas proteicas, mas em alguns casos, a parte catalítica é uma molécula de RNA (Mann et al., 2003).

As endoribonucleases (endo-RNases) iniciam a maturação do RNA e a sua degradação mediante cortes internos, já as exoribonucleases (exo-RNases) o digerem a partir de qualquer extremidade. Como consequência, a ruptura por endo-RNases gera substratos com extremidades 3' e 5' que as exo-RNases posteriormente processam, resultando numa estreita cooperação entre os ambos sistemas. Outras características das ribonucleases incluem sua habilidade para romper fitas simples ou duplas das cadeias do RNA, ou especificamente o RNA em híbridos RNA/DNA (Egli e Martin, 2008; Tadokoro e Kanaya, 2009).

A ação das RNases pode ser progressiva e/ou distributiva (Nogués e Cuchillo, 2001):

quando a enzima se movimenta de um sítio ao seguinte antes de se dissociar do substrato ela é denominada progressiva, e quando o substrato é parcialmente processado e liberado no meio depois da reação inicial, sua ação é denominada distributiva. A distinção entre progressividade e distributividade está baseada fundamentalmente na razão entre as constantes de ruptura e de dissociação, sendo que esta cinética é influenciada por fatores como a força de união enzima/substrato, a concentração de sal, os intermediários na formação dos complexos, etc. Em alguns casos, as condições de reação podem alterar nas RNases as preferências por uma atividade progressiva ou distributiva; já a distinção entre endo e exo-ribonucleases não é alterada de acordo as condições de reação (Nogués e Cuchillo, 2001; Hartung et al., 2010).

Entre as RNases existe uma sobreposição significativa entre as que estão envolvidas na degradação do RNA e aquelas que participam da maturação. Em Bacterias, as endoribonucleases RNases E, G, III e P contribuem para o decaimento do mRNA e para a maturação de RNAs estáveis. Porém, para as exoribonucleases existem diferenças entre as que realizam a degradação ou maturação do RNA. Nas reações de degradação, as enzimas progressivas PNPase, RNase II e RNase R são usadas quase que exclusivamente. Embora estas enzimas possam contribuir para a maturação do RNA, a geração das extremidades maduras geralmente utiliza as enzimas distributivas RNase T, RNase PH e RNase D. Presumivelmente, a ação destas RNases evita a degradação indesejada que pode ocorrer se enzimas progressivas forem usadas. Não obstante, como ocorre a coordenação entre as duas classes de exoribonucleases na maturação do mesmo precursor do RNA continua sem ser explorado (Deutscher, 2006).

Atualmente, são conhecidas cerca de quinze superfamílias de ribonucleases que podem ser agrupadas segundo as suas atividades catalíticas. Exemplos das superfamílias de exoribonucleases são DEED, RNases II/RNB, PDX, DHH e Xrn1/Rat1. Membros das três primeiras superfamílias são encontradas tanto em organismos procariotos como eucariotos. Nas superfamílias de endoribonucleases os membros das superfamílias RNase E/G, RNase III, RNase H cumprem com a característica de estar presentes tanto em organismos procariotos como eucariotos, no entanto os membros das endonucleases PIN, Ferredoxin e Zucchini não cumprem esta característica. Outra superfamília de ribonucleases são as que englobam membros que podem ter tanto atividade exo- como endo-ribonucleolítica. Estas superfamílias são as metalo-

β -lactamase, amplamente distribuídas nos três domínios da vida, e a superfamília HD, com membros identificados somente em organismos procariotos (Arraiano et al., 2013). Uma visão geral das principais famílias de enzimas envolvidas na degradação e processamento dos RNAs celulares são apresentadas na Tabela 1.1.

Tabela 1.1: Principais famílias de enzimas envolvidas na degradação e processamento de RNAs celulares (adaptado de (Arraiano et al., 2013))

Exoribonucleases			
Família	Alguns membros procarióticos	Alguns membros eucarióticos	Comentários
DEDD	RNase D RNase T Om	Pop2/Caf1 Pan2 Rrp6 Rex2-4	* Atividade 3'-5' * Alguns membros tem especificidade de sequência * Membros da família podem ter atividade progressivas ou distributivas
RNase II/ RNB	RNase II RNase R	Dis3 Dss1 Ssd1	* Atividade 3'-5' * Enzimas com atividade progressiva * Distribuídas em todos os domínios da vida
PDX	RNase PH PNPase Exosomos de Archaeas	Rrp41-plantas	* Um o dois domínios PDX por proteínas * Formam estruturas em anel
EEP	-	Ccr4 Ngl1-3	* Atividade 3'-5' * Alguns membros tem especificidade de sequência
DHH	NmA NmB	-	* Chamadas nano RNase
Xm1/Rat1	-	Xm1/Pacman Rat1/Xm2	* Atividade 5'-3' * Relacionados com RNase H
Endoribonucleases			
RNase E/G	RNase E RNase G	RNase E plantas	* Quebra RNA de fita simples
RNase III	RNase III	Rnt1 Drosha Dicer	* Quebra RNA de fita dupla
RNase H	RNase HI, HII e HIII	RNase H1 e H2 Argonautes	* Quebra RNA em complexos RNA/DNA ou RNA-RNA duplex
RegB	RegB	-	* Alvós, preferencialmente, perto as regiões Shine-Delgado
PIN domains	-	Smg6 Dis3(Rrp44)	* Quebra RNA de fita simples * Se pode combinar com outros domínios de nucleases
Ferredoxin-like	Cas nucleases	-	* Altamente específico para CRISPR-RNA
Zucchini	-	Zucchini	* Quebra RNA de fita simples * Envolvido na biossíntese de Piwi-RNA
Ribonucleases com atividade endo- e exoribonucleases			
domínios β -lactamase	RNase J RNase Z	CPSF73 Int9/11 RNaseZ	* Membros da família podem ter atividade exonuclease 5'-3', endonucleases ou as duas atividades. * Muitas vezes como estruturas diméricas (homo ou heterodiméricas) * Membros de família distantes estão envolvidos em reação hidrolítica (destruição de antibióticos, ...)
HD	RNase Y YhaM	-	* Inclui endonuclease e exonucleases 3'-5' * Membros de família dependentes de metal fosfohidrolase que atuam em diversos substratos

1.2 RNases da superfamília das metalo- β -lactamase

Dentro das famílias de RNases reportadas, uma das melhores caracterizadas é a superfamília das metalo- β -lactamase (MBL)(Bebrone, 2007). Geralmente, os membros da família MBL têm um amplo espectro de substratos. Este fato pode se explicado pois seu sítio ativo é acessível um grande número de moléculas; a maioria delas apresenta uma ligação éster e uma carga negativa (L. Aravind and Eugene V. Koonin, 1998).

As proteínas MBL são caracterizadas por cinco motivos de sequências conservadas. Os motivos de sequências conservadas são padrões de histidinas e aspartatos (Figura 1.1). O motivo II tem a sequência consenso HxHxDH (onde x indica qualquer aminoácido), e é a assinatura de toda a família de metalo- β -lactamase (Dominski, 2007). Os motivos facilitam a estrutura terciária específica em duas folhas β flanqueadas por α hélices ($\alpha\beta\beta\alpha$). Conforme determinado por estudos cristalográficos, os resíduos conservados dos cinco motivos formam um centro catalítico contendo tipicamente dois cátions divalentes localizados no fundo de uma ranhura superficial ampla entre as duas folhas β (Figura 1.1C). As MBL ligam principalmente o zinco, porém variantes recombinantes foram isoladas com ferro, manganês e zinco, indicando que o dobramento das MBL permite atividades catalíticas notavelmente diferentes e seletivas a partir da ligação dos cátions (Schilling et al., 2005; Bebrone, 2007). Os membros das MBL são: as glioxalases II, arilsulfatases, fosfodiesterases, proteínas do metabolismo do fosfonato, RNase Z e β -CASP (Fischer et al., 2012). Dentre elas somente as RNase Z e β -CASP, encontradas nos três domínios da vida, estão relacionadas com o metabolismo de ácidos nucleicos (Dominski, 2007).

Os membros da RNase Z são fundamentalmente responsáveis pela maturação das extremidades 3' dos pré-tRNAs que carecem dos nucleotídeos -CCA. Sua atividade expõe o nucleotídeo discriminador, único desemparelhado na extremidade 3' do tRNA, que é usado como primeiro nucleotídeo para iniciar a síntese do triplete -CCA pela nucleotidil transferase do tRNA. O triplete é utilizado como um elemento de identidade para as sintases de aminoacil-tRNA que realizam a aminoacilação. Além deste processo, a RNase Z participa na degradação de mRNA, porém em menor medida (Redko, Li de Lasierra-Gallay e Condon, 2007). Estas enzimas podem ser classificadas em relação ao comprimento de sua sequência em dois grupos: a variante

Figura 1.1: Representação esquemática de proteínas com domínio metalo- β -lactamase relacionadas ao metabolismo dos ácidos nucleicos A) Composição de aminoácidos dos cinco motivos conservados (I-V) e o domínio *exosite*, inserido entre os domínios III e IV, do grupo RNase Z. A baixo a disposição dos aminoácidos do centro catalítico, coordenados aos dois íons zinco. B) Domínios conservados dos membros do grupo β -CASP, posições e composição de aminoácidos dos motivos conservados I-IV e A-C, com o domínio β -CASP inserido como um cassette entre os motivos A e B. Os aminoácidos envolvidos na coordenação dos dois íons de zinco no centro catalítico são mostrados abaixo. C) Estrutura cristalográfica das metalo- β -lactamase contendo dois íons de zinco no centro catalítico (PDB: 1QH3 cadeia Adaptada de (Schilling et al., 2005)).

curta (RNase Z^S, entre 250 e 400 resíduos de aminoácidos) presente em bactérias, arqueias e alguns eucariotos e a variante longa (RNase Z^L, entre 750 e 900) encontradas exclusivamente em eucariotos (Ma et al., 2017; Saoura et al., 2017). A variante curta, RNase Z^S, age como homodímero e além dos motivos I-V, tem um braço flexível (*exosite*) entre os motivos III e IV que permite a interação com os tRNA (Figura 1.1A). A variante longa, RNase Z^L, foi identificada como monômero em solução e propõe-se que surgiu como uma duplicação em tandem da variante curta ou mediante a fusão de dois genes distintos: a variante curta de RNase Z e outro gene com similaridade de sequência a variante curta (Fischer et al., 2012). O domínio C-terminal (CTD) das RNase Z^L compartilha identidade de sequência com as RNase Z^S, mas as sequências do domínio N-terminal (NTD) divergem da família proteica, mesmo assim ele é considerado como um dobramento β -lactamase (Ma et al., 2017; Saoura et al., 2017).

Os membros da família β -CASP englobam enzimas de clivagem de RNA (CPSF73 e RNase J) e reparo do DNA (Artemis, SNM1 e PSO2). A característica dos membros desta família que as distingue de outras MBL é a perda do motivo V; em seu lugar aparece o domínio β -CASP anexado ao domínio MBL e outros três motivos conservados, A-C (Figura 1.1B). O motivo A consiste em um trecho de resíduos hidrofóbicos seguidos de um resíduo ácido. O motivo B contém a histidina conservada, enquanto que o motivo C pode variar entre membros desta família entre os aminoácidos valina e histidina. A estrutura primária dos membros desta família é composta por 460 resíduos, mantendo a característica das MBL na ligação com dois íons zinco. Estas enzimas são capazes de catalisar a degradação endo- e exo-ribonucleolítica. A degradação exoribonucleolítica prossegue na direção 5'-3' e é sensível ao estado de fosforilação da extremidade 5' do transcrito (Dominski, Carpousis e Clouet-d'Orval, 2013).

1.3 Archaea: o terceiro domínio da vida

Archaea é um domínio separado da vida, distinto das Bactérias e Eukarya (Woese, Kandler e Wheelis, 1990; Moissl-Eichinger et al., 2018). Ainda que se assemelham as Bactérias em quanto à sua organização unicelular e falta de núcleo, possuem características únicas, como os componentes da parede e membrana celulares. Por exemplo, as Archaeas não tem peptidoglicanos na sua parede celular, ainda que em algumas espécies se tem detectado a presença de pseudomureína. Já nas estruturas da membrana celular apresentam cadeias de isopreno, L-glicerol e ligações éter, em vez de ligações éster típicas nas membranas celulares de Bactéria e Eukarya (Moissl-Eichinger et al., 2018). Em relação a este último domínio, Archaea compartilham similaridade com as enzimas da maquinaria de processamento molecular, em particular, as enzimas da transcrição, tradução e replicação (Cavicchioli, 2011). Na Tabela 1.2 apresentamos um resumo das principais similaridades e diferenças entre os organismos do domínio Archaea com os organismos dos domínios Bactéria e Eukarya.

Embora esteja bem estabelecido que Archaea e Eukarya compartilham componentes-chave das maquinarias de replicação, transcrição e tradução, resta avaliar em que extensão esse recurso se aplica ao processamento e aos fatores reguladores de RNA. Desses processos, os conhecimentos que se tem em Archaea é muito mais fragmentado comparado com os conheci-

Tabela 1.2: Características apresentadas pelas Bactérias, Archeias e Eucariotos

Característica*	Bactéria	Archaea	Eucarioto
Ligação entre lipídios da membrana celular	Éster	Éter	Éster
Fosfatos dos lipídios da membrana celular	Glicerol-3-fosfato	Glicerol-1-fosfato	Glicerol-3-fosfato
Metabolismo	Bacteriano	Similar ao bacteriano	Eucariótico
Aparato da transcrição	Bacteriano	Similar ao eucariótico	Eucariótico
Fatores de enlogação da tradução	Bacteriano	Similar ao eucariótico	Eucariótico
Núcleo	Não	Não	Sim
Organelas	Não	Não	Sim
Metanogênese	Não	Sim	Não
Patógenos	Sim	Não	Sim

*A ocorrência de determinada característica foi verificada para a maioria dos organismos de cada domínio (a ocorrência não é necessariamente absoluta) - por exemplo, as haloarchaea muitas vezes possuem vesículas de gás, porém a maioria das archaea não possuem organelas (Adaptado de (Cavicchioli, 2011)).

mentos não são de Eukarya como também de Bacteria. Para os principais atores do processamento do RNA -as ribonucleases (RNases)- mais de 30 famílias foram descritas em Bacteria e Eukarya, enquanto apenas poucas foram identificadas em Archaea (Clouet-d'Orval, 2017).

O domínio Archaea é dividido em vários filos: Korarchaeota, Nanoarchaeota, Crenarchaeota e Euryarchaeota (Figura 1.2). Este último filo é o mais extensamente estudado. Ele inclui as archaeas Metanogênicas, que são amplamente distribuídos e desempenham um papel importante nos processos de degradação anóxica de matéria orgânica, e as Halófilas que sobrevivem em concentrações extremas de sal (Moissl-Eichinger et al., 2018).

Recentes descobertas têm mudado a visão sob a biologia de Archaea e revelaram novos traços além da associação com ambientes inóspitos e organismos primitivos. Revelaram, por exemplo, que a partir de linhagens deste domínio emergiram os eucariotos (Eme et al., 2017). Além disso, a sua ampla distribuição em diversos ambientes com o potencial de moldar seu entorno por uma interação tanto com elementos bióticos e abióticos (Moissl-Eichinger et al., 2018), identificadas como componentes substâncias, ou mesmo como espécies chaves, em microbiomas¹ complexos ou acompanhando ao holobionte². Espécies de Euryarchaeota (metanogênicos, halófilos) têm a capacidade de coexistir em microbiomas vegetais, animais e humanos, onde a sintrofia permite que eles prosperem sob estresse por deficiência de energia.

¹O conjunto dos genes encontrados nos microrganismos que vivem no interior ou na superfície de organismos multicelulares complexos – animal e vegetal – conceituam o microbioma.

²O conjunto formado por um organismo multicelular complexo – animal ou vegetal – e todos os seus microrganismos associados e conhecido como holobionte

Figura 1.2: Os três domínios da vida. A divisão da vida neste 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 (Tomado de (Allers e Mevarech, 2005))

Devido a limitações metodológicas, o archaeoma permanece pouco estudado, e muitas questões a respeito de possíveis papéis na patogenicidade, função e interações estruturais com seus hospedeiros e outros microrganismos continuam sendo desconhecidas (Moissl-Eichinger et al., 2018).

1.4 Regulação transcricional e pós-transcricional em Archaeas

Uma única RNA polimerase (RNP) é responsável pela transcrição de todos os RNA em Archaea. RNP está intimamente relacionada com a RNA polimerase II quanto à composição das subunidades, estrutura e uso de fatores de transcrição (Werner e Grohmann, 2011). Os fatores de transcrição (TF) e TATA binding proteins (TBPs) direcionam a iniciação da transcrição em Archaea, de uma maneira similar a como ocorre em Eukarya. Além disso, existem outras proteínas reguladoras que não fazem parte da maquinaria geral da transcrição que tem sido ligadas às respostas fisiológicas e são reconhecidas como reguladores transcricionais, porém semelhantes aos encontrados em Bactéria (Karr, 2014).

O mRNAs de Archaea compartilha características com o mRNA de Bactéria: extre-

midades 5' trifosforiladas, frequentemente numa estrutura policistrónica, não possui introns, nem adição na extremidade 5' de 7-metilguanosina (cap), como tampouco a extensa cauda de poliA na extremidade 3'. Outras características dos mRNA em Archaea é a associação das extremidades 5' trifosfatadas com uma função estabilizadora e, em alguns, mas não em todos, uma cauda heteropolimérica de adenina (A), com uma provável função desestabilizadora na extremidade 3'. Em archaeas halófilas e algumas metanogênicas não foram identificadas modificações pós-transcricionais na extremidade 3' (Evgenieva-Hackenberg e Klug, 2009).

O processo de terminação da transcrição é ainda pouco compreendido em Archaea. Até este momento, sabe-se que: (i) a formação da extremidade 3' do RNA depende de fatores de terminação da transcrição como ocorre em células eucarióticas (Dominski, Carpousis e Clouet-d'Orval, 2013), (ii) o processo é estimulado por sequências de oligo (T₈) (Santangelo et al., 2009) e (iii) a complexidade da sequência de terminação. As sequências terminadoras podem ser estruturadas sem que a estrutura seja essencial para o esse processo, em geral elas encontram-se como regiões consecutivas sugerindo que o processo de terminação é ineficiente ou que as isoformas da região 3'-UTR possam ter um papel na regulação. Além disso, as sequências terminadoras podem estar localizadas no interior das regiões codificadoras de genes que formam parte de operons e o escape da maquinaria de transcrição nestas regiões poderiam permitir a expressão diferencial de genes transcritos de um único promotor (Dar et al., 2016).

Pouco se conhece acerca dos elementos que estabilizam ou desestabilizam os mRNA de Archaea, porém estudos com agentes químicos eficientes para a parada da transcrição tem permitido estimar que o tempo de vida média dos mRNA destes organismos é mais longo que a maioria dos mRNA de Bacteria. Nas bactérias estão na faixa de segundos a minutos e nas arqueias se tem observado que o tempo de vida meia da maioria dos mRNA pode variar entre 5 e 9 min (Evgenieva-Hackenberg e Klug, 2009). Curiosamente, os mRNAs de reguladores traducionais apresentaram meia-vida significativamente mais longa que os mRNAs das correspondentes maquinarias transcricionais e traducionais, sugerindo uma regulação diferencial da expressão gênica no nível pós-transcricional. Esta regulação pode ser útil para uma rápida adaptação às mudanças ambientais e para minimizar as flutuações dos níveis de proteínas na célula (Hundt et al., 2007; Evgenieva-Hackenberg e Klug, 2009).

Avanços na compreensão da estrutura e função do complexo proteico com a capacidade de degradar e processar o RNA têm sido feitos nos últimos anos. Na maioria dos genomas de Archaea estão presentes genes semelhantes aos que formam subunidades do exossomo eucarioto, e juntos ao degradossoma bacteriano incluem proteínas do domínio RNase PH (Büttner, Wenig e Hopfner, 2005; Hundt et al., 2007). Surpreendentemente, nos genomas de *Halobacterium salinarum* NRC-1 e *Methanocaldococcus janaschii* não estão presentes os genes que codificam para subunidades centrais do exossomo, embora o superoperon que o codifica esteja preservado (Hundt et al., 2007; Evguenieva-Hackenberg e Klug, 2009). O resultado sugere que a maquinaria de degradação do mRNA pode ser significativamente diferente (Hundt et al., 2007). Propõe-se que as exoribonucleases tenham papéis chaves na degradação do mRNA. Desta maneira, ortólogos a RNase R na arqueia halofílica *Haloferax volcanii* com função na degradação 3'-5' tem se mostrado essenciais na viabilidade celular (Portnoy e Schuster, 2006); ortólogos a RNase J com atividade endonucleolítica e exo 5'-3' poderiam ser candidatos a degradação do RNA em arqueias metanogênicas (Evguenieva-Hackenberg e Klug, 2009). Contudo, até hoje a contribuição do exossomo das arqueias a vias biológicas específicas ainda é desconhecida e resta determinar se tem algum grau de especialização (Phung et al., 2013).

Em Archaea geralmente os transcritos primários dos tRNA e rRNA são extensivamente processados para produzir as formas maduras e funcionais. Entre os processamentos que ocorrem nos pré-tRNA estão a remoção dos introns e a adição pós-transcricional do triplete -CCA, essencial para o processo de aminoacilação (Redko, Li de Lasierra-Gallay e Condon, 2007). No precursor policistrônico do rRNA, o principal processamento não é só na remoção dos componentes ribossomais 16S, 23S e 5S, mas também dos tRNAs, processando espaçadores internos e distais (Evguenieva-Hackenberg e Klug, 2011). O conjunto destes processamentos são levados a cabo por um grupo de RNases, algumas têm sido identificadas e caracterizadas em passos específicos como RNase Z, RNase P, EndA e Nob1, porém, outras ainda aguardam ser identificadas (Clouet-d'Orval, 2017).

Além dos mRNAs, rRNAs e tRNAs, um amplo grupo de RNA não codificantes (ncRNA) têm sido identificados em Archaea, indicando uma complexa rede de regulação pós-transcricional. Exemplos incluem os snoRNAs (small nucleolar RNAs) (Omer et al., 2000), CRISPRs (Clus-

tered Regularly Interspaced Short Palindromic Repeats) (Clouet-d'Orval, 2017), tRFs (tRNA-derived fragments) (Gebetsberger et al., 2012), TSSaRNAs (Transcription Start Site Associated RNAs) (Zaramela et al., 2014), sotRNA (sense overlapping transcripts) (Gomes-Filho et al., 2015), asRNAs (antisense RNAs) (Almeida et al., 2019) e circRNA (circular RNAs) (Danan et al., 2012). Os caminhos específicos que ditam a biogênese, processamento e degradação desses diversos tipos de RNA têm sido associados em algumas etapas a enzimas ribonucleolíticas. No entanto, as vias e controladores desses processamentos estão em estágios iniciais de caracterização (Clouet-D'Orval et al., 2010; Clouet-d'Orval, 2017).

1.5 *Halobacterium salinarum* NRC-1

No domínio Archaea, um organismo facilmente cultivado e manipulável no laboratório é *Halobacterium salinarum* (Ng et al., 2000). *H. salinarum* é um organismo unicelular, naturalmente encontrado em locais cuja concentração de sal é superior a 4.0 M, como salinas, lagos de água salgada ou até mesmo alojados em cristais de sal. Este organismo é um halófilo obrigatório, suas células apresentam formato de bastonete, flagelos e uma coloração que varia entre as tonalidades rosa e laranja devido a presença de pigmentos carotenóides como as bacterioruberinas e a proteínas de membranas como a bacteriorodopsina (Figura 1.3) (Allers e Mevarech, 2005; Facciotti et al., 2010; Oren, 2012; Jehlička e Oren, 2013). É um organismo anaeróbico facultativo, e nas condições ricas em oxigênio, a síntese de ATP ocorre predominantemente via ciclo do ácido tricarboxílico (TCA) e a fosforilação oxidativa (Ng et al., 2000; Talaue et al., 2016). Em contraste, sob condições anaeróbicas, a síntese de ATP pode ocorrer mediante, pelo menos, duas vias: fermentação de arginina (Ruepp e Soppa, 1996; DasSarma et al., 2006; Oren, 2012) e fototrofia (Oesterhelt e Stoeckenius, 1973; Oesterhelt e Krippahl, 1973). Essa troca metabólica é a marca deste organismo, e em alguns casos, envolve múltiplas estruturas celulares (Gonzalez et al., 2009).

Um componente chave no crescimento fototrófico é a bacteriorodopsina, complexo formado por uma proteína, bacterioopsina, e o derivado de carotenóides, retinal. Várias cópias deste complexo são estão inseridas dentro da membrana celular, conferindo-lhe a coloração púrpura com a que são reconhecidas estes organismos. O complexo funciona como uma bomba

Figura 1.3: *Halobacterium salinarum* NRC-1. A- Microscopia eletrônica (11000X) onde é possível observar a forma bastonete e a presença de flagelos. B- Colônias em cultivo sólido. Coloração rosada devida a presença de pigmentos envolvidos na fototrofia. C- Microscopia eletrônica (11000X) onde é possível verificar a presença das vesículas de gás. (Barra horizontal representa 1µm). Fonte: (Marrero-Gutiérrez, J.)

de prótons impulsionada pelos fótons da luz e que em conjunto com ATP sintase de membrana são capazes de produzir ATP (Skulachev, 1976)(Figura 1.4). Outras bombas estão presentes neste organismo também impulsionadas pela luz como a halorodopsina, capaz de internalizar íons de Cl^- com a função de manter a homeostase celular (Schreiner et al., 2015).

H. salinarum tem duas linhagens bem estudadas: NRC-1 e R1. As principais diferenças entre elas está relacionada com o número de elementos extracromossômicos e a inserção de um fragmento de 10kb no cromossomo de NRC-1 (Pfeiffer et al., 2008). Os estudos em nosso grupo de pesquisa são realizados com a linhagem NRC-1.

O genoma de *H. salinarum* NRC-1 é composto por um cromossomo de aproximadamente 2 Mpb, e dois plasmídeos, pNRC100 com ~190 Kpb e o pNRC200 com ~360 Kpb. Estão anotados no genoma deste organismo 2.629 genes, sendo 2042 no cromossomo, 180 no plasmídeo pNRC100 e 318 no plasmídeo pNRC200. Do total de genes anotados 676 (25,7%) são classificados como hipotéticos, sem função conhecida. Os estudos de proteoma permitiram concluir que apresenta proteínas bem-adaptadas ao ambiente de alta salinidade, apresentando em geral, proteínas ácidas, com um pI médio de 5,1 (Ng et al., 2000; Kennedy et al., 2001; Phu et al., 2009). Estudos relacionados à expressão gênica tem permitido caracterizar mecanismos globais de regulação gênica em resposta a diferentes perturbações ambientais (Kaur et al., 2006; Kaur et al., 2010; Phu et al., 2009; Facciotti et al., 2007), estrutura do transcrito (Koide et

Figura 1.4: Acoplamento quimiosmótico entre a energia solar, a bacteriorodopsina e fosforilação por ATP sintase durante a fototrofia em *Halobacterium salinarum*. A parede celular foi omitida (imagem adaptada do trabalho de Darekk2-Own work, CC BY-SA 3.0,<https://commons.wikimedia.org/w/index.php?curid=21874202>).

al., 2009), replicação (MacNeill, 2009) e mecanismos de reparo do DNA (Busch e DiRuggiero, 2010).

Outro aspecto distintivo de *H. salinarum* NRC-1, é o comportamento bifásico da curva de crescimento. A curva apresenta duas fases de aumento na densidade óptica (OD, optical density) a 600 nm, intercaladas por um período com poucas alterações (Koide et al., 2009). A análise do perfil de expressão gênica ao longo da curva de crescimento permitiu a caracterização da resposta transcricional nos diferentes estágios do crescimento. Na primeira fase exponencial, os principais transcritos encontrados estão relacionados à maquinaria de transcrição, tradução, fosforilação oxidativa e transportadores de ferro. Durante a transição para o estado inicial da fase estacionária, os grupos de genes transcritos estão vinculados a química redox. Na fase estacionária os genes expressos estão relacionados principalmente a transportadores ABC, incluindo transportadores de fosfatos, oligos, dipeptídeos, aminoácidos ramificados o que permite uma adaptação a limitação de nutrientes esperada nesta fase. Acompanhando o esgotamento de nutrientes e a queda na disponibilidade de oxigênio na fase estacionária, observa-se a expressão dos genes envolvidos nos processos alternativos de obtenção de energia. Já a segunda fase exponencial está relacionada com a morte das células e liberação das vesículas de gás no meio de cultura, demonstrada pela diminuição de unidades formadoras de colônia nesse período (Lange

et al., 2007; Facciotti et al., 2010).

Analizadas em conjunto, as características e informações sobre *H. salinarum*, permitem que este organismo se tornasse um importante sistema modelo para caracterização de mecanismos globais de regulação gênica.

1.6 Regulação pós-transcricional em arqueias halofílicas

Apesar das informações disponíveis para *H. salinarum* NRC-1, o conhecimento sobre a regulação pós-transcricional é ainda incipiente, assim como o conhecimento relacionado com os principais atores deste processo molecular, RNases.

Em estudos realizados em 2014, foi relatado no genoma de *H. salinarum* NRC-1 há pelo menos um ortólogo para treze diferentes RNases de linhagens procarióticas e eucarióticas. Na construção de mutantes de deleção desses ortólogos foi descoberto um defeito significativo no crescimento de uma linhagem, Δ VNG2099C. Mediante o estudo do seu transcrito foi possível identificar que a RNase VNG2099 é transcricionalmente co-regulada com genes de estado fisiológico anaeróbico, porém agindo sob os transcritos do estado aeróbico. A atuação desta RNase específica permite uma eficiente transição entre essas condições fisiológicas e resulta em uma redução da taxa de crescimento na linhagem nocaute de seu gene. Neste trabalho, foram construídas com sucessos outras três linhagens nocaute de ortólogos de RNases (Wurtmann et al., 2014).

Uma das linhagens nocaute construídas foi o do gene VNG_RS05855. Esta linhagem, não mostrou um defeito fenotípico significativo sob condições de crescimento de referência em relação ao crescimento da linhagem selvagem. No entanto, foi sugerido que seu possível impacto no crescimento seja em condições específicas (Wurtmann et al., 2014). No RefSeq a sequência do gene VNG_RS05855 está anotada como uma metalo-hidrolase com dobramentos metalo- β -lactamase (Pruitt, Tatusova e Maglott, 2007). As RNases que se caracterizam por ter estes dobramentos estão associadas com o processamento dos pré-tRNAs ou propriamente na degradação do mRNA (Redko, Li de Lasier-Gallay e Condon, 2007; Dominski, Carpousis e Clouet-d'Orval, 2013). Ambos processos são extremamente importantes e poucos compreendidos em *H. salinarum* NRC-1; ainda assim, não há estudos funcionais que caracterizam ao gene

VNG_RS05855.

Além da RNase VNG2099 somente outros dois genes que codificam enzimas ribonucleolíticas foram estudados em *H. salinarum* NRC-1: (1) gene VNG_RS01040, ortólogo da RNase H tipo I de *Escherichia coli* e *Bacillus subtilis*, para o que foi comprovado sua função na remoção do iniciador de RNA nos fragmentos de Okazaki durante a replicação de DNA (Ohtani et al., 2004) e (2) o gene VNG_RS06850, fator de iniciação 5A, homólogo ao fator eucariótico, que participa na ligação e renovação do mRNA (Wagner e Klug, 2007). Outros estudos têm sido realizados para determinar a presença das subunidades do complexo multiproteico capaz de degradar diferentes tipos de RNA, exossomo, e a vida média dos mRNA. Os resultados obtidos permitiram mostrar a ausência das subunidades do exossomo no genoma de *H. salinarum* NRC-1 e a vida média global de seus RNA consideravelmente mais longa do que os RNA de *E. coli* e *B. subtilis*, sendo os transcritos relacionados ao metabolismo energético, transcrição e tradução os que apresentaram maior estabilidade (Hundt et al., 2007).

Os conhecimentos sobre as RNases e o metabolismo do RNA são mais abundantes e variados em *Haloferax volcanii*, organismo do mesmo classe de *H. salinarum* NRC-1 (Halobacteria). A atividade de RNase P, uma endonuclease responsável por remoção de sequências líderes nos precursores de tRNA, já foi demonstrada neste organismo (Nieuwlandt, Haas e Daniels, 1991), assim como a essencialidade da RNase R, apresentando funções similares ao exossomo (Portnoy e Schuster, 2006; Matos et al., 2012). Além disso, foi realizada a primeira caracterização e purificação da ribonuclease que processa a extremidade 3' dos pré-tRNAs, a RNase Z (Schierling et al., 2002). Posteriormente, a especificidade do substrato desta enzima foi estendida a extremidade 5' do rRNA 5S pela semelhança deste fragmento com as estruturas de tRNA (Hölzle et al., 2008). Em 2012, Fischer e colaboradores caracterizaram outro membro da família proteica metalo- β -lactamase, sugerindo a sua participação na osmotolerância e osmoproteção (Fischer et al., 2012). Outro membro estudado desta família proteica foi aCPSF1, com domínio de ribonucleases β -CASP intimamente relacionada ao CPSF73 eucariótico (Cleavage Polyadenylation Specific Factor). Esta proteína foi identificada como componente interativo do proteassoma, sugerindo uma conexão entre processamento de RNA e degradação de proteínas (Chavarria et al., 2014).

Além das ribonucleases citadas aqui para *H. salinarum* NRC-1 e *H. volcanii*, Clouet-D'Orval propôs a existência de outras RNases como EndA, Nob1, aRNase J e aCPSF2, as duas últimas com dobramentos metalo- β -lactamase (Clouet-D'Orval et al., 2010; Clouet-d'Orval, 2017). Esta proposição foi baseada em análises da distribuição filogenética construída a partir do alinhamento de 81 grupos ortólogos de sequências proteicas conservadas em 131 genomas de arqueias completamente sequenciados e com suposta atividade ribonucleolítica. A função de EndA está associada com o *splicing* dos pré-tRNA pela ruptura endonucleolítica; Nob1 com a maturação da extremidade 3' do pré-rRNA da subunidade ribossômica menor; aRNase J e aCPSF2 associadas com o processamento exo-ribonucleotídeo na direção 5'-3', atividade reduzida em substratos de RNA com extremidade 5' trisfosfatados (Clouet-d'Orval, 2017).

O conhecimento dos mecanismos que controlam a qualidade e processamento de RNA em *H. salinarum* NRC-1 em relação ao resto das Haloarchaeas é mais fragmentado e escasso. Compreender os papéis únicos e compartilhados das RNases, ferramentas precisas para o processamento do RNA e frequentemente essenciais na viabilidade celular, é um desafio primordial na construção do conhecimento das vias de regulação pós-transcricional da expressão gênica nesse organismo. Assim, determinar o impacto do gene VNG_RS05855 nos processos da regulação gênica e na estabilidade dos mRNA, não só representa um avanço no conhecimento em *H. salinarum* NRC-1, também pode contribuir ao entendimento de mecanismos pós-transcricionais que permanecem desconhecidos nos terceiro domínio da vida.

2 Objetivos

2.1 Geral

O presente estudo tem por objetivo caracterizar o gene VNG_RS05855 de *Halobacterium salinarum* NRC-1, provável RNase que carece de estudos funcionais (Wurtmann et al., 2014), mediante a exploração de dados públicos, ensaios fenotípicos, análise do transcrito e testes da atividade in vitro de sua proteína heteróloga.

2.2 Específicos

- Caracterizar as sequências do gene VNG_RS05855 e nível de expressão transcricional em distintos contextos genéticos e ambientais com o uso de ferramentas in silico e com dados públicos de *microarray* e *tiling array*.
- Identificar fenótipos específicos associados à linhagem nocaute do gene VNG_RS05855 em diferentes condições ambientais.
- Avaliar o efeito da ausência do gene VNG_RS05855 no transcrito de *H. salinarum* NRC-1.
- Identificar a possível conservação de sequências, ou estruturas secundárias, nos RNAs que possam ser alvo do produto proteico do gene VNG_RS05855, mediante a utilização de metodologias experimentais e computacionais.
- Obter a proteína heteróloga codificada pelo gene VNG_RS05855 e avaliar sua atividade in vitro.

3 Materiais e Métodos

3.1 Banco de dados

Os bancos de dados empregados para as análises da sequência VNG_RS05855 de *H. salinarum* NRC-1 foram do NCBI (National Center for Biotechnology Information): EntreZ-Gene (Maglott et al., 2011), que registra os genomas completamente sequenciados com o objetivo de processar informações específicas dos genes mediante da integração de análises, curadoria e processamento de sequências nucleotídicas; Banco de Domínios Conservados (Conserved Domain Database-CDD), cuja versão mais recente (Marchler-Bauer et al., 2017) contém modelos de domínios proteicos obtidos do Pfam (Finn et al., 2016), SMART (Letunic, Doerks e Bork, 2015), COG (Tatusov, 2001), TIGRFAMS (Haft et al., 2013) e cluster de proteínas (Klimke et al., 2009). Além disso, foram utilizados com as configurações padrões o programa MUSCLE (Edgar, 2004) para obter o alinhamento de múltiplas sequências, inspecionados usando o programa JalView v02 (Waterhouse et al., 2009). A menção a outros bancos de dados empregados será feita na seção onde são utilizados.

3.2 Dados públicos de microarray e tiling array

Os dados públicos utilizados neste trabalho foram ensaios de microarrays e tiling arrays realizados em *H. salinarum* NRC-1 numa variedade de condições ambientais e genéticas (Tabela 3.1). A coleção dos resultados destes experimentos estão disponíveis no banco de dados GEO (Barrett et al., 2013).

Tabela 3.1: Experimentos de *microarray* e *tiling array* de *H. salinarum* NRC-1 utilizados neste trabalho

Condição	Acesso GEO	Referência
Curva de crescimento	GSE12923	Koide et al., 2009
Mudanças na salinidade extracelular	GSE53544	Beer et al., 2013
Mutantes proteínas reguladoras	GSE61975	Plaisier et al., 2014
Mutante de RNase (VNG_RS08130)	GSE45988	Wurtmann et al., 2014
Mudanças de oxigênio		não publicado

O conjunto destes resultados foram utilizados para estudar os níveis de expressão transcricional relativa do gene VNG_RS05855, de genes que codificam proteínas com atividade ribo-

nucleolítica experimentalmente descrita em *H. salinarum* NRC-1 e de genes que estão anotados no RefSeq como essa possível função sem demonstração experimental. Além disso, utilizamos com esses dados estudamos a correlação com a utilização do coeficiente de correlação de Spearman (ρ). A correlação foi considerada forte se os valores modulares do coeficiente de correlação (ρ) estavam entre 0.8 e 1 (Spearman, 1904; Hauke e Kossowski, 2011; Rebekic et al., 2015). O intuito do experimento é sugerir a possível ação coordenada do grupo de genes anotados com a função ribonucleolítica num contexto de mudanças ambientais e genéticas.

3.3 Linhagens e condições de cultivo

As linhagens de *H. salinarum* NRC-1 utilizadas neste estudo foram construídas no Laboratório do Professor Nitin Baliga (*Institute for Systems Biology* – ISB) e gentilmente cedido a nosso laboratório. Os nocautes foram obtidas pelo método *pop-in-pop-out*, que se baseia no processo de recombinação homóloga entre o gene alvo e um fragmento de DNA que contém as regiões flangeadoras do gene alvo a ser deletado (Kaur et al., 2006; Peck, DasSarma e Krebs, 2000). As características das linhagens utilizadas neste projeto: nomes, genótipos e aplicações são resumidas na Tabela 3.2. Como controle do experimento é usada a linhagem Δ ura3, que foi a linhagem precursora para nocautear o gene de interesse VNG_RS05855 (Wurtmann et al., 2014).

Tabela 3.2: Linhagens de *H. salinarum* utilizadas nesta pesquisa

Nomes	Genótipo	Aplicação
Δ ura3 (controle)	Δ VNG_RS06485	RNA-seq
		Ensaio fenotípico
		Testes de atividade
Δ VNG_RS05855 (nocaute)	Duplo nocaute Δ VNG_RS06485 Δ VNG_RS05855	RNA-seq
		Ensaio fenotípico
		Testes de atividade

Denominamos como condição referência de cultivo: temperatura de incubação de 37°C e agitação de 125 rpm em meio CM (NaCl–250 g/L, MgSO₄·7H₂O–20 g/L, Na·Citrate–3 g/L, KCl–2 g/L, e peptona 10 g/L) suplementado com uracila (1 mg/mL), na presença de luz. No cultivo, o primeiro passo foi cultivar pré-inóculos até uma densidade óptica a 600nm (OD_{600nm}) de 0.5, que foram posteriormente diluídos para uma OD_{600nm} ~ 0.05 em volumes que variaram de 200 μ L a 40 mL. A partir destas diluições foram realizadas as medições de absorbâncias

(OD_{600nm}) no decorrer do tempo como medida do crescimento das linhagens.

3.4 Ensaios fenotípicos

A caracterização fenotípica da linhagem nocaute do gene VNG_RS05855 (Δ VNG_RS05855) utilizando a linhagem Δ ura3 como controle incluiu medições de absorbância em diferentes comprimentos de onda, contagens celulares, medições do pH, microscopia eletrônica de transmissão, espectroscopia RAMAN e ensaios de viabilidade celular. Cada uma das metodologias utilizadas será detalhada a seguir.

3.4.1 Curvas de crescimento

A partir das diluições a OD_{600nm} $\sim$ 0.05 dos pré-inóculos das linhagens controle e nocaute foram realizadas medições de absorbâncias (OD_{600nm}) como medida de crescimento (curvas de crescimento) ao longo do tempo (0, 17, 41, 65 e 95 horas). As condições de cultivo de referência foram utilizadas, com a adição de agentes estressantes em diferentes concentrações: ZnSO₄·7H₂O (0.02, 0.05, 0.1, 0.15, 0.20 mM); CuSO₄·5H₂O (0.2, 0.4, 0.8, 1.0, 1.5 mM); FeSO₄·7H₂O (1.0, 2.0, 4.0, 5.0, 6.0 mM) (KAUR et al., 2006), 1,1'-dimetil-4,4'-bipiridina-dicloreto (PQT) (0.125, 0.250, 0.500, 0.750, 1.00 mM) (Kaur et al., 2006; Kaur et al., 2010). Estas condições foram testadas num volume final da cultura de 200 μ L em placas de 96 poços, empregando três réplicas biológicas, cada uma com três réplicas técnicas. Outras condições estressantes foram testadas, em volume de cultura de 40 mL, com igual número de réplicas biológicas: MnSO₄·H₂O (0.8, 1.5 mM) (Webb et al., 2013), CoSO₄·7H₂O (0.2, 0.5 mM) (Woodson, Peck e Krebs, 2003), NaAsO₂ (0.1, 0.2, 0.5 mM) (Gejiao Wang, Sean P. Kennedy, Sabeena Fasiludeen, Christopher Rensing e DasSarma, 2006), H₂O₂ (5, 25 mM) (Kaur et al., 2010) e duas modificações do meio CM, diminuindo a quantidades de sais: meio A com NaCl–215 g/L, MgSO₄·7H₂O–17.2 g/L, Na-Citrate–2.58 g/L, KCl–1.72 g/L, e peptona 10 g/L e meio B com NaCl–180 g/L, MgSO₄·7H₂O–14.4 g/L, Na-Citrate–2.16 g/L, KCl–1.44 g/L, e peptona 10 g/L. O meio A apresenta uma maior concentração de sal que o meio B, com uma variação na porcentagem de sais em relação ao meio CM de 86% e 72% , respectivamente.

3.4.2 Determinação da densidade celular

Entre os métodos mais empregados para estimar a densidade celular está a contagem em câmara de Neubauer (Dyall-Smith, 2008). Para este experimento foram empregadas três réplicas biológicas para as linhagens controle e nocaute, cada uma com uma réplica experimental. Partimos da diluição de pré-inóculos a $OD_{600nm} \sim 0.05$ para iniciar o crescimento. Todas as réplicas foram cultivadas, na condições referência, num volume final de 40 mL em meio CM. Às 17, 41, 65 e 95 horas do crescimento foram coletados 10 μ L de cada amostra. Cada alíquota foi colocada de maneira independente na câmara de Neubauer, previamente posicionada no microscópio ótico, sem uso de corantes nem agentes fixadores para manter a integridade celular inalterada. Utilizando a lente objetiva 40X, foi feita em forma de zigue-zague a contagem de células nos quadrados menores da câmara Neubauer, excluindo das contagens as células que ficavam nas linhas separadoras. Com a expressão $N \times 400 \times 10.000$ células/mL (Dyall-Smith, 2008), estimamos a densidade celular, onde N é o número médio de células presentes nos quadrados menores. A contagem celular em câmara de Neubauer têm um erro de ~20% na estimativa da densidade celular (Bastidas, 2011).

3.4.3 Medições da absorbância em diferentes comprimentos de onda

A partir de três réplicas biológicas por linhagens crescidas num volume de 40 mL de meio CM nas condições de referência, foram realizadas durante o crescimento medições de absorbância a 532, 600, 700 e 785 nm. Realizar medições em diferentes comprimentos de onda permite determinar a contribuição de moléculas fotossensíveis as variações de absorbâncias, como a bacteriorodopsina. A bacteriorodopsina é uma proteína cujo pico de absorção amplo, a 568 nm, pode contribuir para a absorção a 532 e 600 nm, mas não a 700 nem a 785 nm (Facciotti et al., 2010). O aparelho utilizado para realizar as medições de absorbância nos distintos comprimentos de onda foi o Double Mean Spectrophotometer Shimadus.

3.4.4 Medições de pH

Foram crescidas três réplicas biológicas das linhagens controle (Δ ura3) e nocaute (Δ VNG_RS05855), cada uma delas com uma réplica experimental, em 250 mL de meio CM nas condições referência do crescimento. Às 17, 41, 65 e 95 horas de crescimento foram coletadas alíquotas de 15 mL em tubos Falcon. Das alíquotas foram feitas medições de pH empregando o eletrodo de pH, HI2221-HANNA. As mesmas alíquotas foram centrifugadas a 5.000 x g durante 10 min a 4°C e novas medições de pH foram feitas no sobrenadante.

Às 17, 41, 65 e 95 horas foram feitas duas novas diluições das linhagens em meio CM e duas com solução basal de sal (NaCl–250 g/L, MgSO₄·7H₂O–20 g/L, Na·Citrate–3 g/L, KCl–2 g/L), numa relação 1:5 em volume final de 15 mL. Das duas diluições, uma foi mantida na presença da luz e a outra no escuro. As medições de pH se realizaram nas diluições mantidas na luz ou no escuro após 0, 20, 40, 60, 80 e 110 minutos. O tratamento estatístico dos dados foi realizado verificando a distribuição normal das medições de pH com o teste de Shapiro (Wilk e M.B., 2005). Caso a distribuição fosse normal, foi aplicado o teste *t-student* (Boneau, 1960) e em caso contrário o teste de “*rank*” de Wilcoxon (Wilcoxon, 1945).

3.4.5 Espectroscopia RAMAN

Da curva de crescimento das linhagens foram retiradas amostras de 2 mL às 17, 41, 65 e 95 horas de crescimento, centrifugadas a 12.000 x g durante 5 minutos a 4°C. O sobrenadante foi descartado e o pellet foi seco por speedvac durante 20 minutos. O pellet foi estocado a -80°C até a realização da espectroscopia Raman. Os espectros Raman foram obtidos no Laboratório Nacional de Luz Síncroton (LNLS) com a colaboração do Dr. Douglas Galante.

Brevemente, um espectro Raman é a luz que espalha a amostra após de ser excitada com radiação monocromática no visível ou na região do infravermelho próximo. A luz que emerge tem mudanças na sua frequência, ocasionadas, entre outras causas, pela interação com as moléculas que conformam as amostras. A luz promove nelas transições vibracionais que deformam a sua nuvem eletrônica, afetando a polarizabilidade. Cada molécula vai ter um conjunto de valores de frequência característicos da sua estrutura e das ligações químicas que a formam.

Então, os picos nos espectros Raman representam a intensidade relativa nas mudanças destas frequências após a excitação (entendidas como deslocamento/alongamento no comprimento de onda -cm^{-1}) e estão relacionadas com as vibrações de estruturas moleculares (Marshall et al., 2006; Leuko et al., 2007; Fendrihan, Musso e Stan-Lotter, 2009). Excitando as amostras no visível conseguimos o aumento na intensidade dos picos Raman na quarta potência, porém são criadas sinais de fundo que dificultam a interpretação dos resultados. Por esse, motivo é recomendável, também fazer a excitação das amostras no infravermelho próximo, para diminuir o efeito perturbador da fluorescência de fundo (Fendrihan, Musso e Stan-Lotter, 2009).

Os espectros foram obtidos com o aparelho da empresa Renishaw, excitando as amostras com comprimentos no visível (532 nm) e no infravermelho próximo (785 nm); os dados foram coletados depois de um tempo de exposição de 1 segundo, 30 acumulações e uma laser power/% de 50. Os intervalos de varredura foram de $600\text{-}1600\text{ cm}^{-1}$, correspondentes com o intervalo de visualização dos carotenóides, bioassinatura empregada para as haloarchaeas (Jehlička e Oren, 2013; Jehlička, Edwards e Oren, 2014). Utilizamos o programa WiRE v4.3 para processar os espectros: aplicar o baseline e realizar a normalização. O baseline é um método para corrigir a linha base de espectros pela subtração de uma constante, eliminando as regiões dos espectros que não tem muita informação. A normalização foi realizada utilizando como parâmetro os valores da intensidade RAMAN, para que cada espectro obtido varie entre o limite inferior de 0 e o limite superior 1. O arquivo foi salvo como .txt. Os espectros normalizados foram trabalhados com o pacote de R “ChemoSpec” (Hanson, 2014) para a visualização.

3.4.6 Microscopia eletrônica de transmissão

Todas as réplicas (três réplicas biológicas e cada uma com a sua réplica experimental) das diferentes linhagens utilizadas neste experimento foram cultivadas num volume final de 40 mL em meio CM nas condições de referência. Às 17, 41, 65 e 95 horas do crescimento foram coletados 200 μL da cultura, mantidos durante 5 minutos em grades de Formvar, antes de examiná-las no microscópio eletrônico Philips CM120 fazendo corridas de 10 kV e ampliações 11.000 X.

3.4.7 Ensaios de sobrevivência

Para avaliar a viabilidade celular, 5 mL da cultura controle e da linhagem Δ VNG_RS05855 em fase exponencial ($DO_{600nm} \sim 0.5-0.6$) foram aliquotadas em tubos Falcon. As alíquotas foram recolhidas às 1, 3 e 6 horas após a adição de agentes estressantes em concentrações semi-letais ($CuSO_4 \cdot 5H_2O$ (1.25 mM); $NaAsO_2$ (0.3 mM) e PQT (4.0 mM)) (Kaur et al., 2006; Gejiao Wang, Sean P. Kennedy, Sabeena Fasiludeen, Christopher Rensing e DasSarma, 2006; Kaur et al., 2010) e colocado em gotas de 10 μ L em placas de CM em diluições $10^{-2}, 10^{-3}, 10^{-4}$. As placas foram incubadas por 7 dias a 37°C e fotografadas. Com as fotografias se realizaram as análises do número de unidades formadoras de colônia (UFC) com o software ImageJ (Schneider, Rasband e Eliceiri, 2012). Os resultados correspondem a três réplicas biológicas para cada linhagem, e para cada uma delas, duas réplicas experimentais por diluição.

3.5 Estudo do transcrito

Para caracterização do gene VNG_RS05855 analisamos as mudanças no transcrito ocasionada pela ausência do mesmo. Nesta seção iremos descrever as metodologias moleculares e o conjunto de ferramentas computacionais que empregamos nas análises dos resultados destes ensaios.

3.5.1 Extração do RNA total e pequenos RNAs

Foram cultivadas três réplicas das linhagens controle e nocaute (Δ VNG_RS05855) nas condições referência, alíquotas de 2 mL foram coletadas e centrifugadas a 12.000 x g por 5 minutos a 4°C, o sobrenadante foi descartado e precipitados formados mantidos a -80°C. A partir deles se realizou a extração do RNA.

Para a extração do RNA foi utilizado o kit mirVana miRNA Isolation (Ambion) que permite a separação seletiva do RNA total e RNAs pequenos. Para a extração do RNA total foram utilizados dois precipitados por extração e para a extração do RNA pequeno foram empregados três precipitados. Após do processo de lise celular foi realizada a remoção dos

resíduos celulares mediante a extração com fenol/clorofórmio. Depois da extração orgânica adicionou-se 1.25X volume de etanol 100% na extração do RNA total. No caso do RNA pequeno adicionou-se primeiramente 1/3 do etanol absoluto e se passou pela coluna de extração do kit mirVana e ao volume recuperado adicionou-se 2/3 do volume de etanol 100%. A etapa de adição de etanol é crítica para determinar o grau de afinidade dos RNAs pelas colunas de separação, dessa forma a adição de etanol nas quantidades descritas promovem a ligação seletiva dos RNAs. Após passagem na coluna, os RNAs foram lavados com os tampões fornecidos no kit e então eluídos em água DEPC a 95°C e quantificados em espectrofotômetro.

3.5.2 Tratamento com DNases

Para remoção do DNA residual das amostras foi utilizado 50 µg de RNA total e todo o volume recuperado do RNA pequeno após da purificação. Para cada reação foram adicionados 10 µL de Turbo DNase (Thermo Scientific), 15 µL de Turbo DNase 10x tampão e 0.5 µL de Ribolock (Thermo Scientific) para um volume final de 150 µL. A reação foi incubada a 37°C por 45 minutos. Em seguida, foram adicionados 150 µL de fenol ácido as amostras, que foram vortexadas e prontamente centrifugadas a 10.000 x g por 5 minutos. As fases aquosas das amostras foram transferidas para novos microtubos onde se adicionou 2.5x o volume de etanol 100% e 1/10 do volume de NaOAc 3 M. As reações foram incubadas a -80°C overnight. As amostras foram centrifugadas a 14.000 x g, 4°C por 30 minutos, os sobrenadantes desprezados, o precipitado lavado com 500 µL de etanol 70%, em seguida as amostras foram centrifugadas por 5 minutos a 10.000 x g. As amostras foram secas a 65°C por 10 minutos, e ressuspensas em 40 µL de água MiliQ DEPC.

Para confirmação da correta degradação dos contaminantes de DNA nas amostras foram realizadas reações de PCR utilizando 10 µM dos pares de primers 19/20 e 63/64 (Tabela 3.3) que amplificam regiões conservadas no plasmídeo pNRC200 e no cromossomo de *H. salinarum* NRC-1, respectivamente, além de 6.25 µL do GoTaq (Promega), 5% de DMSO (Kapa Biosystems). As condições da PCR foram de 95°C por 15 minutos, seguidos de 30 ciclos de 95°C por 15 segundos, 58°C por 15 segundos e 72°C por 30 segundos. Como controle positivo foi utilizado DNA genômico de *H. salinarum* NRC-1. O produto da PCR foi submetido à ele-

troforese em gel de agarose 1% em tampão de corrida TBE 1X, utilizando o marcador de peso molecular GeneRuler 1Kb plus.

Tabela 3.3: *Primers* utilizados neste trabalho

Primers	Sequências	Aplicação
19-fwd	5'- AACCACCGTGACCCGGAAGC - 3'	confirmação da correta degradação dos contaminantes de DNA
20-rev	5'- GCCGATCCGATGGGAGTGACT - 3'	confirmação da correta degradação dos contaminantes de DNA
63-fwd	5'- GTGAGGGACGAAAGCTAGGG - 3'	confirmação da correta degradação dos contaminantes de DNA
64-rev	5'- AAGGGTACTGCTGGCAATTA - 3'	confirmação da correta degradação dos contaminantes de DNA
VNG_RS05855-fwd	5'- ctgacacatatgATGCAACTGACGTTTCTCGGC - 3'	Amplificação do gene VNG_RS05855
VNG_RS05855-rev	5'- tcgatcgatccCTAGAGCTCGACGCCGAGAC - 3'	Amplificação do gene VNG_RS05855

3.5.3 Enriquecimento dos RNAs primários

O enriquecimento de RNAs primários foi realizado nas amostras de RNA total com a enzima TEX (Terminator 5'-Phosphate-Dependent Exonuclease) capaz de degradar os RNAs monofosfatados na extremidade 5', ou seja, RNAs que passaram por alguma etapa de processamento. O resultado deste tratamento são amostras enriquecidas com RNA trifosfatado (primário). Para o tratamento foram utilizados 10 µg do RNA total divididas em duas alíquotas de 5 µg cada. Uma delas foi tratada com 1U da enzima TEX (Epicentre TER51020) para cada 1 µg de RNA (TEX+) enquanto a segunda alíquota foi incubada apenas na presença do tampão da enzima (TEX-). As reações foram incubadas a 30⁰C por 60 minutos e logo após foi realizada a purificação com o kit RNeasy MinElute Cleanup (Qiagen). Para verificar se houve degradação dos RNAs processados (RNAs com extremos 5' monofosfatados) 1µL das amostras TEX+ e TEX- foi analisada em eletroforese em gel de agarose 1% com tampão de corrida TBE. A confirmação da ausência de bandas correspondentes do RNA ribossomal (rRNA) na electroforese indica a depleção dos RNA processados nas amostras TEX+, o contrário é esperado para amostras TEX-. Posteriormente, as amostras foram tratadas com a enzima TAP (Tobacco Acid Pyrophosphatase) (Epicentre T19100) na concentração de 1U para cada 1 µg de RNA a 37⁰C por 30 minutos, para liberar a molécula de mRNA com apenas um grupo fosfato na extremidade 5, e purificadas novamente com o kit RNeasy MinElute Cleanup (Qiagen)

3.5.3.1 Software para a identificação em larga escala de inícios da transcrição

O sequenciamento em larga escala das amostras enriquecidas e não enriquecidas do

RNA primários é conhecido como dRNA-seq (differential RNA-seq) e como vai ser realizado o sequenciamento irá ser detalhado mais adiante. No entanto, a comparação do produto de sequenciamento de amostras tratadas com a enzima TEX (TEX+), que degrada preferencialmente o RNA processado ou monofosfato no seu extremo 5', enriquecendo as amostras com RNA primários, com o produto de sequenciamento das amostras não tratadas com a enzima (TEX-) formadas pela coleção de conjunto dos RNA, primários e processados, têm permitido identificar, de uma forma global, os sítios que marcam os inícios de transcrição numa variedade organismos procariotos (Sharma et al., 2010; Ten-Caten et al., 2018; Fröhlich, Förstner e Gitai, 2018)(Figura 3.1).

Figura 3.1: A técnica de dRNA-seq consiste no tratamento da amostra de RNA com a enzima 5' Terminator Exonuclease (TEX). Essa enzima degrada RNAs que apresentam apenas um grupo fosfato na extremidade 5', ou seja, transcritos que tenham passado por alguma etapa de processamento. Com isso a amostra é enriquecida para RNAs primários. A análise do sequenciamento desses RNAs, em conjunto com dados de uma biblioteca controle sem enriquecimento, permite a identificação das posições de início de transcrição (TSS). (Adaptado de (Sharma e Vogel, 2014)).

Um software bem estabelecido como ferramenta para a detecção automatizada dos sítios que marcam os inícios da transcrição (TEX + > TEX -) é TSSAR stand-alone (Amman et al., 2014; Ten-Caten et al., 2018). Brevemente, a ferramenta analisa o enriquecimento de contagem de início de leituras (reads) nas bibliotecas preparadas com as amostras TEX+ em comparação com a biblioteca das amostras TEX-, levando em consideração o nível de expressão da região na qual a posição analisada se localiza. Essa comparação é realizada assumindo-se que

a contagem de inícios de reads segue uma distribuição de Poisson com parâmetro λ definido pela média das contagens na região. Dado que a diferença entre duas distribuições de Poisson pode ser modelada de acordo com uma distribuição de Skellam (Skellam, 1946), todas as posições podem ser avaliadas verificando se a diferença de contagem entre as bibliotecas está ou não de acordo com a distribuição esperada para a região. Caso essa diferença seja maior do que o esperado a posição é determinada como um sítio de início da transcrição (TSS-transcription start site) (Amman et al., 2014). A aplicabilidade deste software no análises dos dados do sequenciamento do dRNA-seq tem sido estendida para determinar outras posições relevantes nos genomas de distintos organismos, como sítios de processamento do transcritos (Xu et al. 2015; Qi et al. 2017; Amr em preparação).

3.5.4 Depleção de RNA ribossomal

Alíquotas de 5 μ g do RNA total purificado de ambas linhagens foram selecionadas para realizar a depleção do RNA ribossomal (rRNA) utilizando o kit Ribo-Zero rRNA Removal (Illumina). No processo cada amostra é incubada durante 5 minutos a 68⁰C com probes que hibridizam especificamente com o rRNA. Posteriormente, probes-rRNA são incubados com beads magnéticos durante 5 minutos a 50⁰C. Os beads são capazes de interagir com os probes-rRNA formando um aglomerado terciário que é removido após incubação no suporte magnético por 2 minutos a temperatura ambiente. Depois do processo de depleção são recuperados aproximadamente 8% da quantidade de RNA inicial, visto que entre 82% a 90% do RNA total é composto por rRNA (Blazewicz et al., 2013).

3.5.5 Fragmentação do RNA depletado

Alíquotas de 500 ng do RNA total depletado do RNA ribossomal foram submetidas ao processo de fragmentação, segundo o protocolo Illumina. Cada amostra é diluída em tampão de fragmentação 10X (1 M KOAc, 300 mM Mg(OAc)₂, 400 mM Tris-OAc) até o volume final de 20 μ L e incubadas a 94⁰C durante 4 minutos. Nesta condição são obtidos fragmentos que variam entre 150-350 pb. Depois deste processo as amostras são purificadas com o kit RNeasy MinElute Cleanup (Qiagen) recuperando 16 μ L. O processo de fragmentação compromete a integridade das extremidades 5' dos RNA, razão pela qual as amostras devem ser tratadas. O

primeiro passo é a incubação dos 16 μ L com 1 μ L de fosfatase antartica (Epicentre), 2 μ L de seu tampão 10X e 1 μ L de RNaseOut (Qiagen) no termociclador durante 30 minutos a 37°C, e posteriormente 5 minutos a 65°C. Em seguida, as amostras são misturadas com 2 μ L de T4 polinucleotideo quinase (NEB), 5 μ L do tampão 10X da fosfatase, 5 μ L de ATP a 10mM e 1 μ L de RNaseOut (ThermoFisher), levando-as ao termociclador para a incubação durante 60 minutos a 37°C. Por último, as amostras foram purificadas empregando kit RNeasy MinElute Cleanup (Qiagen).

3.5.6 Preparação das bibliotecas de RNA-seq

Para a preparação das bibliotecas temos como material de partida para as linhagens controle e nocaute amostras de RNA total enriquecido com RNA primário (depletado de RNA processados), RNA total depletado de rRNA sem fragmentar e RNA total fragmentado depois da depleção do rRNA. A quantificação para iniciar a preparação das bibliotecas foi feita com kit Quant-iT™ RiboGreen® RNA Assay (Invitrogen). Por se tratar de uma quantificação mediante marcação por fluorescência específica para moléculas de RNA, esta metodologia minimiza a interferência por nucleotídeos livres ou outros contaminantes, fornecendo uma quantificação do RNA mais precisa.

Para a preparação das bibliotecas utilizamos dois kits: o TruSeq Small RNA Sample Preparation (Illumina) e TruSeq mRNA Stranded RNA Sample Preparation (Illumina). O emprego de dois kits se justifica na eficiência para conservar a especificidade de fita, sendo no kit TruSeq Small melhor conservada esta informação, segundo a experiência de trabalhos prévios em nosso laboratório.

3.5.6.1 TruSeq Small

Com o kit TruSeq Small foram preparadas as bibliotecas do RNA total depletado do rRNA sem fragmentar, do RNA total enriquecido com RNA primário e de RNA total fragmentado depois da depleção do rRNA.

Das amostras de RNA total depletado do rRNA foram preparadas quatro bibliotecas, duas bibliotecas para as réplicas biológicas da linhagem controle (Δ ura3) e outras duas para as réplicas da linhagem nocaute (Δ VNG_RS05855). Das amostras de RNA total enriquecido

Tabela 3.4: Resumo do estudo das técnicas de RNA-seq para estudar o transcrito do Δ VNG_RS05855

Preparo de bibliotecas	Amostras	Linhagens				Sequenciamento
		Δura		ΔVNG_RS05855		
TruSeq Small	RNA total depletado do rRNA (sem fragmentar)	Réplica1	Réplica2	Réplica 1	Réplica 2	MiSeq Reagent v2 50 cycles (Paired-end)
	RNA total enriquecido com RNA primário (dRNA-seq)	TEX +	TEX -	TEX +	TEX -	MiSeq Reagent v2 300 cycles (Paired-end)
	RNA total depletado do rRNA (fragmentado)	Réplica 1		Réplica 1		MiSeq Reagent v2 300 cycles (Paired-end)
TruSeq mRNA Stranded	RNA total depletado de rRNA (fragmentado)	Réplica 1		Réplica 1		MiSeq Reagent v2 300 cycles (Paired-end)

com RNA primário também foram preparadas quatro bibliotecas, duas para a linhagem Δ ura3 (uma para a amostra tratada com TEX (TEX+) e outra para amostra não tratada (TEX-)) e duas bibliotecas para a linhagem Δ VNG_RS05855 (uma TEX+ e a outra TEX-), sem réplicas biológicas. Foram também preparadas duas bibliotecas com RNA total fragmentado depois da depleção do rRNA para cada linhagem, também sem réplicas (Tabela 3.4).

A construção das bibliotecas se iniciou com 500 ng de RNA. A primeira etapa da preparação consiste na ligação de adaptadores 3' e 5' às moléculas de RNA, seguida de uma etapa de transcrição reversa para síntese de cDNA e uma posterior amplificação por PCR que adiciona um index único a cada biblioteca, além de aumentar o número de cópias das moléculas que serão posteriormente sequenciadas. Após estes procedimentos as amostras foram corridas em gel 6% Novex TBE PAGE (Invitrogen) e a porção logo abaixo do marcador de 145 pb e acima do marcador de 500 pb foi excisada para purificação, selecionando fragmentos até o tamanho de 600 pb por se tratar do tamanho máximo em que as moléculas são capazes de formar ligações em bridge para obtenção dos clusters no flowcell. As porções do gel correspondente as amostras foram fragmentadas e eluídos em H₂O MilliQ overnight, sendo purificadas em tubos com filtros de 5 μ m.

3.5.6.2 TruSeq mRNA Stranded

Com o kit TruSeq mRNA Stranded foram preparadas duas bibliotecas com as amostras do RNA total depletado de rRNA das linhagens controle (Δ ura3) e a linhagem nocaute (Δ VNG_RS05855) (Tabela 3.4).

As construções das bibliotecas iniciou-se com 900 ng do RNA. O primeiro processo foi

a fragmentação do RNA meio de cátions divalentes e temperatura elevada empregando o tampão de fragmentação recomendado pelo fabricante (Illumina). Posteriormente, é realizada a síntese do cDNA utilizando transcriptase reverse e primers aleatórios. Nessa etapa foram acrescentados 0.12 µL de actinomicina D (1 µg/µL) para aumentar a especificidade das sínteses e prevenir a geração de fragmentos de cDNA espúrios. A especificidade da cadeia das amostras é conseguida substituindo dTTP por dUTP no processo de síntese de segunda cadeia, o que interrompe a amplificação subsequente, uma vez que a polimerase utilizada no ensaio não prossegue com a síntese além deste nucleotídeo. O produto final deste passo são moléculas de cDNA de cadeia dupla com extremidades blunt. A adição subsequente de uma base A nas extremidades das cadeias de cDNA impede que os fragmentos de extremidade blunt se liguem um ao outro durante a reação de ligação de adaptadores. Em seguida, ocorre uma reação de ligação, pela qual os adaptadores de indexação específicos são ligados a cada amostra nas extremidades dos fragmentos de DNA, preparando-os para hibridação na flowcell. Os produtos são purificados e enriquecidos com PCR para criar a biblioteca de cDNA final.

3.5.7 Quantificação das bibliotecas

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. Dessa forma, podemos aferir a quantidade exata de moléculas aptas para sequenciamento e ajustar o volume necessário de cada biblioteca com o intuito de obter quantidades similares de sequências e por conseguinte quantidades similares de reads ao final do processo. Após quantificação, a concentração de cada biblioteca foi ajustada para 1.7 nM.

3.5.8 Sequenciamento das bibliotecas

O sequenciamento das bibliotecas foram realizados com dois kits diferentes (Tabela 3.4). O kit MiSeq Reagent v2 50 cycles Paired-end, empregado para sequenciar as quatro bibliotecas de RNA totais preparadas a partir do protocolo TruSeq Small. O kit MiSeq Reagent v2 300 cycles Paired-end foi empregado para as quatro bibliotecas de RNA total enriquecido com RNA primários, as duas bibliotecas de RNA total fragmentado depois da depleção de

rRNA preparadas a partir do protocolo TruSeq Small, além das duas bibliotecas de RNA total depletado do rRNA preparadas com o protocolo de TruSeq Stranded. O processo de sequenciamento foi executado conforme as instruções do fabricante. Na Tabela 3.4 estão resumidos o RNA utilizado, as diferentes bibliotecas, o número de réplicas de cada linhagens e o kit de sequenciamento para cada uma delas.

3.5.9 Processamento dos dados de sequenciamento

O processamento dos dados foi realizado primeiramente removendo os adaptadores dos reads obtidos com o emprego da ferramenta Trimmomatic (Bolger, Lohse e Usadel, 2014). Posteriormente, os reads foram alinhados ao genoma referência de *H. salinarum* NRC-1 (NC_002607, NC_002608, NC_001869) com o software HISAT (Kim, Langmead e Salzberg, 2016). Com o intuito de aumentar o número de reads alinhados no genoma os parâmetros do programa foram mudados, permitindo que os reads alinharam em mais de um lugar, impossibilitando neles a formação de gaps e spliced. Após esta etapa se empregou o software Multipapper Resolution (Kahles, Behr e Rätsch, 2015) para a seleção do total de reads alinhados em mais de um lugar só aqueles onde seu alinhamento no genoma minimize a variância local, diminuindo a inflação das contagens e permitindo a exploração de uma maior quantidade de regiões no genoma. Com a ferramenta Deeptools (Ramírez et al., 2014) são gerados as contagens de reads por posição do genoma e com a função summarizeOverlaps do pacote GenomicAlignments (Hervé Pagès, Valerie Obenchain, 2018) a matriz de contagens de reads por genes, para este resultado foi necessário dar como input o arquivo .GFF que contém as anotações dos genes do organismo em estudo (Ten-Caten et al., 2018). Este protocolo para o processamento dos dados de sequenciamento está disponível em github (<https://github.com/alanlorenzetti/frtc/>).

Os arquivos com as contagens de reads por posição do genoma de cada biblioteca foram processados para a visualização na ferramenta Gaggles Genome Browser-GGB (Bare, 2010).

Já com a matriz de contagem de reads por genes foi possível realizar o cálculo da expressão diferencial utilizando o pacote DESeq2 (Michael I Love e Anders, 2014). Como este pacote foram analisados os resultados do sequenciamento das quatro bibliotecas de RNA totais preparadas a partir do TruSeq Small e sequenciadas MiSeq Reagent v2 50 cycles Paired-End, sendo a única biblioteca que contém réplicas biológicas das linhagens controle e nocaute. Foram

considerados genes diferencialmente expressos na relação nocaute versus controle aqueles que apresentaram $\log_2\text{FoldChangel} > 1$ e p-value ajustado < 0.05 . Com as matrizes de contagens de reads por genes nas restantes bibliotecas foram feitos análises dos gráficos de MA-plot. Estes gráficos representam o $\log_2$ da média das contagens de reads normalizadas (A) versus o $\log_2$ da razão da contagem de reads (M) em cada gene das linhagens na relação nocaute vs controle. Identificamos aqueles com expressão diferencial se os valores de $|\text{M}(\log_2\text{FC})| > 1$ e $\text{A}(\log_2 \text{ da média de contagens}) > 5$ (Pobre e Arraiano, 2015).

Da lista de proteínas codificadas pelos genes diferencialmente expressos obtidas pelo DESeq2 e MA-plot foi realizada, mediante os termos Gene Ontology (GO), a análise do enriquecimento de processos biológicos. Primeiro foram anotadas, segundo os termos GO, as sequências proteicas de *H. salinarum* NRC-1 reportadas no RefSeq (Pruitt, Tatusova e Maglott, 2007) com ferramenta Sma3s (Sequence massive annotator using 3 modules) que se auxilia do banco de dados Uniprot, realizando anotações que são altamente sensíveis e específicas (Casimiro-Soriguer, Muñoz-Mérida e Pérez-Pulido, 2017). O próximo passo foi identificar os termos enriquecidos nossa lista de interesse com o pacote TopGO aplicando o teste exato de Fisher (Alexa, Rahnenführer e Lengauer, 2006).

3.6 Exploração de cobertura de reads obtida do sequenciamento de amostras de RNA total depletado de rRNA sem fragmentar

A cobertura de reads do RNA total depletado de rRNA sem fragmentar preparada com o kit TruSeq Small (Tabela 3.4) tem duas potencialidades i) o RNA não está fragmentado, ii) o protocolo de sequenciamento conserva a informação da especificidade da fita. Ambas potencialidades, a nosso ver, podem auxiliar na identificação de possíveis sítios de processamento da suposta atividade ribonucleolítica do produto proteico do VNG_RS05855. Para isso, primeiro reconhecemos as posições de início e fim dos reads sequenciados em cada linhagens. Posteriormente, numa mesma linhagem determinamos as posições adjacentes no genoma que apresentem um acúmulo de reads. Definimos o acúmulo de reads como significativo se na posição i do genoma o número de reads é maior que 5, e maior que o número de reads na posição adjacente ($i+1$). Na posição $i+1$, o número de reads tem que ser maior que o 50% do

número de reads na posição i . Esta consideração permite evitar desbalanceamentos entre o número de reads das posições que podem ter como causa um processamento a posteriori do sítio de reconhecimento. Comparando as posições que cumpram com estas premissas entre linhagens, previmos, que a posição que esteja presente na linhagem controle (Δura3) e ausente na linhagem nocaute ($\Delta\text{VNG_RS05855}$) representa uma potencial posição do processamento da atividade do produto proteico do gene VNG_RS05855 (Figura 3.2).

Figura 3.2: Representação esquemática da identificação de possíveis sítios de processamento do produto proteico do gene VNG_RS05855 com dados do sequenciamento de amostras de RNA total depletadas de rRNA preparadas com o protocolo TruSeq-Small. Morro azul representam a cobertura de reads no transcrito processado (vermelho) o não processado (preto) pela ação da suposta ribonucleases (“Pac-Man”)

3.7 Obtenção da proteína recombinante do gene VNG_RS05855 e ensaios de atividade

Para a analisar a função in vitro da proteína codificada no gene VNG_RS05855 de *H. salinarum* NRC-1, geramos a proteína recombinante mediante a expressão heteróloga em *Escherichia coli*. Nesta seção descrevemos o conjunto de procedimentos experimentais utilizados para este objetivo.

3.7.1 Expressão e purificação da proteína recombinante

Foram utilizadas a linhagem Δura3 de *H. salinarum* NRC-1, cultivada nas condições

de referência, e as linhagens DH5 α e BL21 de *E. coli*, crescida aerobicamente na temperatura de 37°C em meio LB suplementado com ampicilina (100 μ g/mL). A sequência do gene VNG_RS05855, obtida do RefSeq-NCBI, foi a base para a construção dos primers forward (VNG_RS05855-fwd) e reverse (VNG_RS05855-rev) (Tabela 3.3) que permitiram a sua amplificação a partir do DNA genômico da linhagem controle. A amplificação foi realizada via PCR de colônia utilizando 10 μ M deste par de primers, 6.25 μ L de GoTaq (Promega), 5% de DMSO (Kapa Biosystems). As condições da PCR foram de 95°C por 15 minutos, seguidos de 30 ciclos de 95°C por 15 segundos, 58°C por 15 segundos e 72°C por 30 segundos. O produto de PCR foi submetido a eletroforese em gel de agarose 1% em tampão de corrida TBE 1X, utilizando o marcador de peso molecular GeneRuler 1Kb plus. A banda correspondente a amplificação do gene VNG_RS05855 foi cortada do gel da eletroforeses e realizada a purificação do fragmento segundo o protocolo Wizard®SV Gel and PCR Clean-Up Sytem (Promega), seguindo os passos do fabricante.

Os fragmentos purificados resultantes da amplificação por PCR foram clonados no vetor pET15b. Para isso tanto o inserto como o vetor foram previamente digeridos com as enzimas NdeI e BamHI. A digestão se realizou de maneira independente para o inserto e o vetor, utilizando 5 μ L de cada, 10 U de cada enzima, 2 μ L do tampão EcoRI no volume final de 20 μ L. As reações foram incubadas por 3 horas a 37°C e foram purificadas com o Wizard® Plus SV Minipreps DNA Purification System (Promega) seguindo as instruções do fabricante. As concentrações do inserto e vetor foram determinadas utilizando o aparelho NanoDrop One/One (Thermo) e com auxílio do NEBicalculator (<https://nebiocalculator.neb.com/#!/ligation>). A ligação foi feita com a relação 1:3 de inserto:vetor, utilizando 200 U da enzima T4 ligase (NEB), 2 μ L do tampão da enzima no volume final de reação de 20 μ L, que se manteve durante toda a noite a 16°C.

Com o produto da ligação foram transformadas células competentes de *E. coli* DH5 α por eletroporação. Utilizamos o aparelho Gene Pulser Xcell Electroporation System, seguindo as instruções do fabricante para a eletroporação de bactérias. Após este processo as células foram reconstituídas em 600 μ L de meio LB e colocadas em eppendorf de 1.5 mL deixando 1 hora sob agitação a 37°C. Depois deste tempo, todo o volume foi vertido em placas com

meio LB + Ampicilina (100 µg/mL) (marcador de seleção) e incubadas a 37°C durante 24 horas. As placas foram retiradas da estufa e um grupo de colônias selecionadas para testar quais tinham incorporado o vetor com o inserto. Os testes foram realizados mediante PCR de colônias usando os mesmos pares de primers (VNG_RS05855-fwd e VNG_RS05855-rev, Tabela 3.3) e as mesmas condições do PCR utilizadas na amplificação do gene VNG_RS05855 a partir do DNA genômico de linhagem controle (Δ ura3). Das colônias confirmadas com a construção vetor mais inserto foi feito um pré-inóculo em meio LB+Ampicilina. O crescimento foi quantificado mediante medições de absorbância a 600nm. Quando estes valores atingiram 0.5 foi feita a purificação da construção (vetor + inserto) com o kit Wizard® Plus SV Minipreps DNA Purification System (Promega) seguindo as instruções do fabricante.

Com a construção purificada, a linhagem BL21 (DE3) de *E. coli* foi transformada por eletroporação utilizando o aparelho Gene Pulser Xcell Electroporation System, seguindo as instruções do fabricante para a eletroporação de bactérias. Essa cepa foi cultivada em 500 mL de meio LB com ampicilina até OD_{600nm} 0.6-0.8, quando então foi adicionado isopropyl β -D-1-thiogalactopyranoside (IPTG) na concentração final de 1 mM, por 2 horas. Após este tempo, a cultura foi centrifugada a 14.000 x g por 5 minutos a 4°C. O pellet bacteriano formado foi res-suspendido em tampão de lise (tampão fosfato de sódio 50 mM pH 7,4, NaCl 300 mM, imidazol 50 mM e PMSF 1 mM) e sonicado (10 pulsos com amplitude de 75% por 20 segundos com intervalos de 59 segundos). Em seguida, foi adicionado estreptomicina (10% v/v) à solução, que foi então filtrada (poro de 0.45 µm) e purificada por cromatografia líquida em coluna de níquel Ni-NTA Superflow (Qiagen), seguindo as orientações do fabricante. As frações contendo a proteína de interesse foram concentradas utilizando Vivaspin® 6 Concentrator (Sartorius Stedim Biotech), e o tampão de eluição foi trocado em colunas Disposable PD 10 Desalting (GE Healthcare) para o tampão de estocagem (Tris-HCl 20 mM pH 7,4, NaCl 100 mM e Glicerol 5%). A avaliação final do processo de purificação da proteína foi feito resolvendo os extratos totais e as frações obtidas ao longo do processo em gel de poliacrilamida 15% (SDS-PAGE). O tampão de corrida Tris/Gly (25 mM Tris-HCl, 250 mM de Gly, 0.1% SDS p/v), a 200 V durante 45 minutos. A posterior quantificação da proteína recombinante foi feita com a reagente colorimétrico Bradford utilizando o kit da Bio-Rad (Coomassie Brilliant Blue R-250), seguindo as

instruções do fabricante (Bio-Rad).

3.7.2 Ensaios de atividade da proteína recombinante

A atividade da proteína recombinante (rVNG_RS05855) foi testada *in vitro* utilizando RNA total e RNAs pequenos das linhagens nocaute (Δ VNG_RS05855) e controle (Δ ura3). Neste experimento o primeiro passo foi realizar as quantificações do RNA (total e pequeno) com o kit RNA RiboGreen® segundo as instruções do fabricante (RiboGreen), medindo a fluorescência no leitor CytoFluor™ II (Biosistemas Perceptivos).

O ensaio de atividade foi realizado com 40 µg da proteína recombinante, 4 µg do RNA total ou 2 µg do RNA pequeno num volume de reação de 100 µL, empregando como tampão de reação (50 mM MES pH 5.5, 5 mM MgCl₂, 10 mM KCl, 2 mM DTT) o descrito em (Späth et al., 2008). O tempo de reação foi 45 minutos a 37°C. Como controles de atividade foram empregados, além dos substratos mais o tampão de reação sem a proteína recombinante, as RNases A (ThermoFisher), RNase R (Epicentre), RNase S1 (Invitrogen) e RNase V1 (Ambion) testadas com o substrato RNA total. Após do tempo de reação, a atividade das proteínas foi interrompida mediante a extração com fenol/clorofórmio. Este processo foi feito com a adição de 150 µL de fenol ácido as amostras, que foram vortexadas e prontamente centrifugadas a 10.000 x g por 5 minutos. As fases aquosas das amostras foram transferidas para novos microtubos onde se adicionou 2.5x o volume de etanol 100% e 1/10 do volume de NaOAc 3M. As reações foram incubadas a -80°C overnight. As amostras foram centrifugadas a 14.000 x g, 4°C por 30 minutos, os sobrenadantes desprezados, o precipitados lavados com 500 µL de etanol 70%, em seguida as amostras foram centrifugadas por 5 minutos a 10.000 x g. Posteriormente secadas a 65°C por 10 minutos, e ressuspensas em 10 µL de água MiliQ DEPC.

Os produtos das reações, assim como RNA total e pequeno das linhagens controle (Δ ura3) e Δ VNG_RS05855, foram analisadas com o kit Bioanalyzer RNA (Agilent) seguindo as instruções do fabricante (Technologies, 2006; Technologies, 2016). Este kit que permite separar e quantificar mediante eletroforese capilar o RNA, tem entre suas variantes o kit Bioanalyzer RNA 6000 Nano para caracterizar o RNA total e o Bioanalyzer Small RNA que permite separar e quantificar com precisão RNA de tamanho abaixo de 200nt (RNA pequenos) (Technologies, 2006; Technologies, 2016). Os produtos destas eletroforeses foram analisa-

dos com o software Agilent 2100 Expert (Agilent-2100).

4 Resultados

Nesta seção apresentaremos os resultados da caracterização do gene VNG_RS05855 de *H. salinarum* NRC-1. Iniciaremos mostrando as análises da sua sequência de nucleotídeos e aminoácidos, seguido de análises da expressão transcricional em diversas condições ambientais e genéticas a partir de dados públicos de microarray e tiling arrays. Em seguida, mostraremos os resultados de diferentes ensaios fenotípicos comparando a linhagem Δ VNG_RS05855 com a linhagem controle: curvas de crescimento, medições de pH, espectroscopia RAMAN, microscopia eletrônica e ensaios de sobrevivência. Apresentaremos os resultados da avaliação do impacto no transcrito de *H. salinarum* NRC-1 da deleção do gene VNG_RS05855 mediante ensaios de RNAseq. Mostraremos possíveis sítios de processamento dos transcritos identificados mediante análises de dados de sequenciamento de amostras enriquecidas de RNAs com extremidades 5' monofosfatadas. Analisaremos a conservação das estruturas secundárias das sequências onde foram mapeados os sítios de processamento, buscando identificar possíveis sinais de reconhecimento. Por último, apresentaremos os resultados da expressão, purificação e o teste da atividade da proteína recombinante codificada pelo gene VNG_RS05855.

4.1 A sequência proteica do gene VNG_RS05855 conserva os motivos das proteínas metalo- β -lactamase com função ribonucleolítica

O gene VNG_RS05855 está anotado no genoma de *H. salinarum* NRC-1 como uma possível metalo-hidrolase com dobramentos metalo- β -lactamase (MBL). Para identificar genes que estivessem caracterizados experimentalmente e que compartilhassem identidade de sequência com o gene VNG_RS05855, realizamos o alinhamento local da sua sequência de nucleotídeos (BLASTN) na base de dados nucleotídeos do NCBI. Foram identificadas um total de sete sequências, todas de genes de outras arqueias halofílicas, com identidade variando entre 72% e 100% (Tabela 4.1). Até o momento, nenhum destes genes estão caracterizados funcionalmente. Embora os genes com dobramentos MBL estejam anotados na maioria dos genomas de Archaea, muitos deles têm a sua função desconhecida (Fischer et al., 2012). Nesse contexto, pesquisas que abordem a atividade do produto proteico de metalo-hidrolase com dobramentos

Tabela 4.1: Organismos que apresentam genes com maior percentual de identidade de sequências com o gene VNG_RS05855.

Organismos	Genes	% de cobertura	E-value	% de identidade
<i>Halobacterium salinarum</i> R1	OE_3149F	100%	0.0	100%
<i>Halobacterium salinarum</i> strain 91-R6	HBSAL_08040	100%	0.0	100%
<i>Halobacterium</i> sp. DL1	HALDL1_12200	98%	6.0e-176	83%
<i>Halobacterium hubeiense</i> strain J120-1	HHUB_2631	99%	8.0e-140	80%
<i>Halorubrum trapanicum</i>	CPZ01_RS13175	89%	7.0e-066	75%
<i>Halodesulfurarchaeum formicicum</i> strain HSR6	HSR6_0709	73%	2.0e-027	72%
<i>Halodesulfurarchaeum formicicum</i> strain HTSR1	HTSR_0683	73%	7.0e-026	72%

metalo- β -lactamase ganham relevância por serem aspectos poucos explorados.

Utilizamos o banco de dados de domínios conservados (CDD) do NCBI para identificar as proteínas com as quais o gene VNG_RS05855 compartilha dobramentos MBL. As proteínas com descrição ribonuclease Z (RNase Z) foram as principais proteínas identificadas (Tabela 4.2). As proteínas assim descritas estão relacionada fundamentalmente ao processamento dos pré-tRNAs; no entanto em *Haloferax volcanii* foi demonstrado que as RNase Z são capazes de processar precursores do rRNA 5S (Hözl et al., 2008) e em *Escherichia coli*, o envolvimento no processamento de mRNA (Redko, Li de Lasierra-Gallay e Condon, 2007).

As RNases Z pertencem a família proteica das MBL. O dobramento estrutural das MBL foi descrito pela primeira vez para β -lactamases, enzimas hidrolíticas que conferem resistência bacteriana aos antibióticos β -lactâmicos. Entre os membros desta família estão glioxalases II, arilsulfatases, fosfodiesterase e, junto às RNase Z, as β -CASP que são as que com este tipo de dobramento participam no metabolismo dos ácidos nucleicos (Vogel et al., 2005; Fischer et al., 2012).

Tabela 4.2: Dez proteínas com os menores E-value que apresentam os domínios conservados identificados no gene VNG_RS05855.

Proteína	ID	Descrição	Interv. Seq.	E-value
ElaC	COG1234	Ribonuclease BN, tRNA processing enzyme	1-741	4.61e-58
RNaseZ_MBL-fold	cd16272	Ribonuclease Z; MBL-fold metallo-hydrolase domain.	7-534	1.42e-44
RNaseZ_ZIPD-like_MBL-fold	cd07717	Ribonuclease Z, E. coli 3' tRNA-processing endonuclease ZIPD	7-729	1.98e-38
RNase_Z	TIGR02651	ribonuclease Z; Processing of the 3-prime end of tRNA precursors	7-741	4.65e-33
PRK00055	PRK00055	ribonuclease Z; Reviewed	1-741	3.36e-31
RNaseZ_short-form-like_MBL-fold	cd07716	uncharacterized bacterial subgroup of Ribonuclease Z, short form	4-534	1.20e-30
metallo-hydrolase-like_MBL-fold	cd07740	uncharacterized subgroup of the MBL-fold_metallo-hydrolase superfamily	10-567	5.84e-28
metallo-hydrolase-like_MBL-fold	cd07741	uncharacterized subgroup of the MBL-fold_metallo-hydrolase superfamily	7-567	3.33e-21
RNaseZ_ELAC1/ELAC2_MBL-fold	cd07718	Ribonuclease Z ELAC1, C-terminus of ELAC2, and related proteins	7-528	6.83e-15
Lactamase_B_2	pfam12706	Beta-lactamase superfamily domain	88-633	9.87e-15

Para estudar a conservação dos motivos das MBL no produto proteico do gene VNG_RS05855

realizamos o alinhamento múltiplo, incluindo sequências proteicas das RNases Z de organismos procariotos nas que se demonstrou atividade ribonucleolítica, tanto *in vivo* como *in vitro*, (Schierling et al., 2002; Vogel et al., 2005; Schilling et al., 2005) (SCHIERLING et al., 2002; VOGEL; SCHILLING; SPA; et al., 2005), o motivo carboxi-terminal da RNase Z^L de *Saccharomyces cerevisiae* que conserva os resíduos catalíticos (Ma et al., 2017; Saoura et al., 2017), um conjunto de outras metalo-hidrolases com dobramentos MBL que compartilham identidade de sequência com o produto proteico do gene VNG_RS05855, e a sequência proteica do gene VNG_RS08725 de *H. salinarum* NRC-1, anotada como possível RNase Z no nos bancos de dados não-redundantes do NCBI.

Provavelmente a proteína codificada pelo gene VNG_RS08725 é a RNase Z de *H. salinarum* NRC-1, esta proposta é feita com a base na essencialidade da sua função no processamento dos pré-tRNAs deste organismo e ao fato de que tentativas de construir mutantes de deleção deste gene não resultaram em colônias viáveis (Baliga NS, ISB, comunicação pessoal). Sendo assim, como explicar a função de outra proteína, como a codificada no gene VNG_RS05855.

No resultado do alinhamento observa-se os cinco motivos conservados da família MBL e a ausência do exosite em um grupo sequências (Figura 4.1). O exosite, localizado entre os motivos III e IV, está ausente na região carboxi-terminal da RNase Z^L de *S. cerevisiae* e no conjunto de metalo-hidrolases com dobramentos MBL, incluindo VNG_RS05855. A perda do exosite na região carboxi terminal da RNase Z^L de *S. cerevisiae* não afeta o reconhecimento dos pré-tRNA visto que esta função é feita pela sua região amino terminal (não incluída neste alinhamento). No entanto, para o conjunto de metalo-hidrolases com dobramentos MBL, a ausência do exosite compromete o reconhecimento dos pré-tRNAs. Com este achado, não é possível atribuir o produto proteico do gene VNG_RS05855 a possível função da RNase Z no processamento direto dos pré-tRNAs de *H. salinarum* NRC-1.

Figura 4.1: Alinhamento múltiplo da sequência de aminoácidos do gene VNG_RS05855 com proteínas RNase Z caracterizadas em organismos procariotos (no alinhamento os organismos com estas proteínas estão identificadas com a letra Z), com a sequência VNG_RS08725, possível RNase Z de *H. salinarum* NRC-1, com a região carboxi terminal da RNase Z^L de *Saccharomyces cerevisiae* (Sce*) e com proteínas metalo-hidrolases com dobramentos MBL de organismos procariotos com os que compartilha identidade de sequência. As sequências estudadas pertencem a: Bsu, *Bacillus subtilis*; Eco, *Escherichia coli*; Tma, *Thermotoga maritima*; Tac, *Thermoplasma acidophilum*; Pae, *Pyrobaculum aerophilum*; Pfu, *Pyrococcus furiosus*; Mja, *Methanocaldococcus jannaschii*; Hvo, *Haloferax volcanii*; Gau, *Gemmatimonas aurantiaca*; Mma, *Methanosarcina mazei*; Afu, *Archaeoglobus fulgidus*. O alinhamento mostra os cinco motivos característicos da família metalo- β -lactamase (indicados por I-V) e o motivo braço flexível (ou *exosite* - indicado no quadro preto-).

Apesar da proteína codificada pelo gene VNG_RS05855 ter poucas chances de reconhecer pré-tRNAs como alvos, não podemos descartar a sua possível atividade ribonucleolítica. Esta ideia é plausível ao analisar o alinhamento da sequência de aminoácidos do gene VNG_RS05855 com o motivo catalítico carboxi-terminal da RNase Z^L de *S. cerevisiae* (Sce*) e observar que compartilha as características de conservação dos aminoácidos do domínio funcional e ausência do exosite (Saoura et al., 2017; Ma et al., 2017). Outra perspectiva para analisar a possível atividade ribonucleolítica do produto proteico do gene VNG_RS05855 são os relatos da capacidade da RNase Z de processar outros RNAs mRNA (Redko, Li de Lasier-Gallay e Condon, 2007; Hölzle et al., 2008). No processamento destas variantes de RNA não se conhece a função do exosite, visto que o nível de estruturação destas moléculas é diferente.

4.2 Os níveis de expressão relativa das RNases de *H. salinarum* NRC-1 são baixos, segundo as análises nos dados públicos de microarray e tiling array

As análises da sequência do gene VNG_RS05855 sugerem um possível papel ribonucleolítico de seu produto proteico. Está documentado na literatura que as RNases são necessárias para processos de maturação ou degradação do RNA (Evguenieva-Hackenberg e Klug, 2011; Wurtmann et al., 2014; Clouet-d'Orval, 2017), estritamente reguladas durante o crescimento celular e sofrem influência ambiental (Deutscher, 2006; Wurtmann et al., 2014). A ação das RNases pode ser coordenada de forma cooperativa, redundante ou não específica (Clouet-d'Orval, 2017).

Partindo dessas ideias, utilizamos o conjunto de resultados de experimentos de microarray e tiling array de *H. salinarum* NRC-1 para estudar os níveis de expressão relativa de algumas RNases. O intuito desta análise é identificar se genes que codificam possíveis RNases teriam perfis de expressão semelhantes ou distintos em relação ao gene VNG_RS05855, em diversas condições ambientais ou backgrounds genéticos. Estes resultados poderiam auxiliar a identificar possíveis ações de cooperação, redundância das RNases e/ou especificidades de acordo com as condições ambientais.

Avaliamos juntamente com o gene VNG_RS05855, três RNases caracterizadas experimentalmente em *H. salinarum* NRC-1 (RNase H, IF5-A, RNase-VNG2099) e outros dois genes (VNG_RS08725 e VNG_RS02375) com função ribonucleolítica (segundo a base de dados RefSeq do NCBI) demonstrada experimentalmente na maturação dos pré-tRNAs em procariotos (Nieuwlandt, Haas e Daniels, 1991; Schierling et al., 2002). Cinco contextos genéticos e ambientais distintos foram avaliados: a curva crescimento de *H. salinarum* NRC-1 (I), nas linhagens nocaute de proteínas reguladoras (VNG_RS05365 - ativador transcricional- e VNG_RS04845 - repressor transcricional-)(II) e nocaute da RNase-VNG2099 (III), além das mudanças de oxigênio (IV) e de salinidade extracelular (V) (Figura 4.2).

Figura 4.2: Perfis de expressão relativa de genes com possível atividade ribonucleolítica em ensaios de *microarray* e *tiling array* em distintas condições genéticas e ambientais. Os diferentes genes são: A) VNG_RS05855 (em estudo); B) VNG_RS08725 (anotado como RNase Z); C) VNG_RS01040 (anotado como RNase H com atividade demonstrada experimentalmente (Ohtani et al., 2004); D) VNG_RS06850 (anotado como fator de tradução IF-5A com atividade ribonucleolítica demonstrada experimentalmente (Wagner e Klug, 2007); E) VNG_RS08130 (anotado como imine deaminase com atividade ribonucleolítica demonstrada experimentalmente: RNase-VNG2099 (Wurtmann et al., 2014)); F) VNG_RS02375 (anotado como RNase P). As condições representadas são: I) Curva de crescimento (Koide et al., 2009); II) Nocaute de proteínas reguladoras (Lo et al., 2014); III) Nocaute de RNase-VNG2099 (VNG_RS08130) (Wurtmann et al., 2014); IV) Mudanças de oxigênio (não publicados); V) Mudanças nas condições de salinidade (Beer, Wurtmann e Baliga, 2014).

Observa-se que para o gene VNG_RS05855 os níveis de expressão relativa são bai-

xos em todas as condições experimentais testadas (Figura 4.2 A). Padrão semelhante é observado para os genes VNG_RS08725 (RNase Z), VNG_RS01040 (RNase H) e VNG_RS08130 (RNase-VNG2099) (Figura 4.2 B, C e E), o que poderia indicar um papel constitutivo na sua expressão. Para os genes VNG_RS08725 (RNase Z), VNG_RS06850 (IF5-A) e VNG_RS02375 (RNase P) sempre observamos valores negativos de expressão relativa na curva de crescimento (Figura 4.2 B_I, D_I e F_I). Estes níveis de expressão, distintos dos observados para o gene VNG_RS05855, podem estar relacionados com o controle durante o crescimento. IF5-A, homólogo ao mesmo fator em eucariotos, participa no contínuo turnover do mRNA durante o crescimento de *H. salinarum* NRC-1 (Wagner e Klug, 2007), enquanto a RNase Z e RNase P tem como principal função em organismos procariotos a maturação dos pré-tRNAs (Vogel et al., 2005; Redko, Li de Lasier-Gallay e Condon, 2007).

O conjunto de perfis de expressão mostram que, embora esses genes apresentem baixa expressão diferencial, as diferentes RNases mudam os perfis de expressão nas distintas condições.

A interconexão das atividades das RNases, seja na maturação ou degradação do RNA, é um fenômeno conservado nos três domínios da vida (Chen, Penny e Collins, 2011; Pobre e Arraiano, 2015; Vogel et al., 2005). Mesmo que os níveis de expressão sejam baixos, correlações no perfil de expressão podem ser informativos para identificar RNases com perfis semelhantes ou antagônicos. Assim, ao invés de estudar somente RNases com atividade demonstrada, utilizamos todos os 18 ortólogos de RNases anotados no genoma de *H. salinarum* NRC-1, nas mesmas cinco condições ambientais avaliadas anteriormente, e calculamos o coeficiente de correlação Spearman (ρ).

No decorrer da curva de crescimento e nas mudanças de salinidade (condição I e V), não conseguimos identificar transcritos de possíveis RNases correlacionados ao produto transcricional do gene VNG_RS05855 ($|\rho| > 0.8$) (Figura 4.3).

Já nas linhagens nocaute de proteínas reguladoras (condição II) e linhagem nocaute da RNase-VNG2099 (condição III) (Figura 4.3), onde a perda das funções destas proteínas deve alterar a expressão de transcritos (Lo et al., 2014; Wurtmann et al., 2014), observa-se a correlação positiva ($\rho > 0.8$) entre o transcrito do gene VNG_RS05855 e um grupo de transcritos

que codificam possíveis RNases de *H. salinarum* NRC-1. Nas condições II e III a correlação se estabelece com os mesmos dois transcritos, os dos genes VNG_RS10275 (RNA_binding protein) e VNG_RS01560 (aCPSF1), o que pode sugerir que as adaptações às mudanças genéticas impliquem mecanismos similares de regulação pós-transcricional para o controle das alterações nos níveis de transcritos. Junto ao transcrito do gene VNG_RS10275, que segundo análises de sua sequência de aminoácidos na base de domínios conservados do NCBI (CDD) evidenciou motivos similares com a RNase E/G (E-value= 2.46e-18), na condição II, o transcrito do gene em estudo também se correlaciona ao transcrito do VNG_RS04485 (RNase J). A RNase E de *E.coli* exibe a redução da atividade *in vitro* frente a substratos 5'-trifosfatados, sendo dependente do processamento endonucleolítico desta extremidade (Deutscher, 2006; Clouet-D'Orval et al., 2010). A correlação positiva entre transcritos de prováveis RNases poderia indicar alguma colaboração entre seus produtos proteicos para garantir a função. Na condição III, a correlação com o transcrito VNG_RS10370 (RNase R) também poderia ser outro exemplo desta colaboração. A RNase R atua sobre RNA estruturados mediante o reconhecimento de curtas extremidades 3' livres (Hossain, Malhotra e Deutscher, 2016), a correlação com os transcritos de outras prováveis RNases seria para promover o prévio processamento de RNA estruturados com extremidades 3' longas.

Figura 4.3: Correlação da expressão dos genes anotados como RNases em dados de *microarray* e *tiling array* em distintas condições genéticas e ambientais usando o coeficiente de Spearman (ρ). As condições: I) Curva de crescimento (Koide et al., 2009); II) Nocaute de proteínas reguladoras (Lo et al., 2014); III) Nocaute de RNase-VNG2099 (VNG_RS08130) (Wurtmann et al., 2014); IV) Mudanças de tensão de oxigênio (não publicados); V) Mudanças nas condições de salinidade (Beer, Wurtmann e Baliga, 2014). Os genes com asterisco indicam atividade que seus produtos proteicos são RNase com atividade experimentalmente demonstrada

A condição em que se observa maior quantidade de transcritos de possíveis RNases correlacionados positivamente ao transcrito do gene VNG_RS05855 é na condição IV (Figura 4.3) que avalia as mudanças na tensão de oxigênio. Treze dos dezoito transcritos associados a possíveis ribonucleases estão positivamente relacionados ($\rho > 0.8$: VNG_RS01560, VNG_RS02375, VNG_RS05000, VNG_RS05010, VNG_RS05865, VNG_RS06595, VNG_RS06850, VNG_RS07685, VNG_RS08130, VNG_RS08615, VNG_RS08725, VNG_RS09830, VNG_RS10275),

sugerindo papel das RNases nessas condições. O papel-chave das RNases no controle de processos relacionados ao metabolismo de oxigênio em *H. salinarum* NRC-1 como os de fermentação de arginina, respiração anaeróbica (TMAO e DMSO, como aceptores finais dos elétrons), citocromos oxidases e produção das vesículas de gás já foi sugerida (Baliga et al., 2002; Kaur et al., 2010; DasSarma et al., 2012), nossos resultados corroboram esta sugestão e adicionamos VNG_RS05855 neste conjunto.

Ao analisar as condições onde se estabelecem correlações negativas ($\rho < -0.8$), observamos que só ocorrem nas mudanças genéticas, II e III (Figura 4.3) com cinco genes: VNG_RS01040 (RNase H), VNG_RS07685 (RNase HII), VNG_RS02375 (RNase P), VNG_RS08615 (RNase tRNA splicing endonuclease), VNG_RS08725 (RNase Z). Deles, três: RNase P, tRNA-splicing e RNase Z, codificam proteínas que estão relacionados a processos de maturação dos tRNAs e rRNAs (Redko, Li de Lasierra-Gallay e Condon, 2007; Hölzle et al., 2008). A correlação negativa com transcritos podem indicar que, perante alterações genéticas o produto proteico do gene VNG_RS05855 não auxilia na maturação dos RNA e sim na sua degradação, posto que nestas condições supomos permita esta função a melhor adaptação.

Uma visão integrada do conjunto destes resultados indica que a expressão do VNG_RS05855 pode estar relacionada com modificações nos RNAs que garantam o posterior processamento por outras RNases. Estas modificações podem não ser indispensáveis para a atividade das outras RNases, mas a ação coordenada poderia influenciar com precisão nos níveis do RNA na célula e consequentemente na adaptação às mudanças ambientais ou genéticas. Neste último caso, poderia favorecer os processos de degradação acima dos de maturação dos RNA.

4.3 Ensaios fenotípicos

Um passo importante na caracterização da funcionalidade de VNG_RS05855 foi a obtenção da sua linhagem nocaute. A linhagem foi gentilmente cedida pelo laboratório do Prof. Nitin S. Baliga do Institute for Systems Biology (ISB-Seattle, EUA) e permitiu o desenvolvimento de uma diversidade de experimentos para identificar possíveis fenótipos representativos de mudanças em processos celulares/moleculares originados pela ausência do gene VNG_RS05855.

4.3.1 A linhagem nocaute do gene VNG_RS05855 é viável e apresenta crescimento similar à linhagem controle

Nós iniciamos a caracterização fenotípica observando as variações da OD_{600nm} no decorrer do tempo das linhagens nocaute (Δ VNG_RS05855) e controle (Δ ura3). Durante o crescimento de *H. salinarum* NRC-1 os incrementos da OD_{600nm} nas etapas finais podem não ser devidos somente ao crescimento celular mas sofrer influência da liberação ao meio extracelular de vesículas de gás ou a produção de compostos fotossensíveis como a bacteriorodopsina, ambos fatores capazes de incrementar a dispersão da luz (Facciotti et al., 2010). Por essa razão, as análises das variações da OD_{600nm} das linhagens foram complementadas com medições da densidade celular.

A caracterização fenotípica preliminar realizada por Wurtmann et al, 2015 mostraram diferenças significativas entre as linhagens nocaute e controle nas fases finais do crescimento (Wurtmann et al., 2014). Porém, nossos dados não mostram diferenças significativas entre as linhagens em nenhuma das fases do crescimento (Figura 4.4), não conseguindo reproduzir os resultados obtidos por Wurtman e colaboradores. Vale ressaltar que as condições experimentais na realização da curva de crescimento foram diferentes: Wurtmann et al. utilizaram 200 μ L de meio em um sistema automatizado (BioscreenC) para avaliar o crescimento das linhagens, nós realizamos o crescimento em 40 mL de meio. Um volume menor de meio poderia representar um incremento na probabilidade de evaporação e consequente aumento na precipitação de sais e afetar proporcionalmente as medições de absorbâncias. Estas diferenças experimentais poderiam ser as que originem as diferenças fenotípicas que observamos em nossos dados.

Verificamos a diminuição da densidade celular nas fases finais do crescimento em nossos dados experimentais (Figura 4.4). Este comportamento é típico do crescimento de *H. salinarum* NRC-1 e coincide com o reportado por Facciotti e col, permitindo explicar que os incrementos da OD_{600nm} não são proporcionais ao aumento da densidade celular (Facciotti et al., 2010). Contudo, não conseguimos identificar diferenças significativas nas densidades celulares entre as linhagens durante o crescimento, ainda quando é fácil apreciar que a média da densidade celular na linhagem nocaute é menor que na linhagem controle sem ser uma diferença significativa.

Figura 4.4: Média das variações da absorbância a 600nm (OD_{600nm}) (eixo da esquerda corresponde com o diagrama de linhas) e da densidade celular (eixo da direita corresponde com o diagrama de barras) no decorrer do tempo do conjunto de réplicas das linhagens controle ($\Delta ura3$ -vermelho) e nocaute ($\Delta VNG_RS05855$ -preto). As linhas verticais representam o desvio padrão dos dados.

Wurtman et al sugerem que o produto proteico do gene VNG_RS05855 tenha um papel em condições específicas do crescimento (Wurtmann et al., 2014), com base na relação das RNases com mecanismos de regulação pós-transcricionais que influenciam a adaptação dos organismos a diversas condições ambientais (Deutscher, 2015). Entre as diversas condições ambientais que têm sido estudadas em *H. salinarum* NRC-1 estão os estresses causados por diversos metais de transição e mudanças nas condições da salinidade do meio (Kaur et al., 2006; Kaur et al., 2010; Beer, Wurtmann e Baliga, 2014). Para verificar a influência desses estresses na linhagem nocaute, realizamos ensaios de curvas de crescimento sob diferentes estresses ambientais.

4.3.2 Sob estresses o crescimento da linhagem nocaute apresenta sutis diferenças em relação à linhagem controle

Foram realizadas curvas de crescimento utilizando diferentes concentrações de cinco sais de metais de transição (Zn^{2+} , Cu^{2+} , Fe^{2+} , Mn^{2+} , Co^{2+}), um metalóide ($NaAsO_2$), dois

agentes oxidantes (PQT e H_2O_2) e duas variações de sais do meio CM.

4.3.2.1 Metais de transição

Os metais de transição avaliados têm um vínculo direto com mecanismos relacionados a parada do crescimento, eliminação de íons tóxicos, aumento do turnover de proteínas, reparo de DNA e modulação dos transportadores ativos de íons em *H. salinarum* NRC-1. Nessas condições, foram caracterizadas discrepâncias entre os níveis de mRNA e proteínas como uma possível consequência de efeitos na regulação pós-transcricional (Kaur et al., 2006).

Nossos resultados mostraram pequenas diferenças nas $\text{OD}_{600\text{nm}}$ entre as linhagens nas fases finais da curva de crescimento na presença de alguns metais. Na presença de CuSO_4 0.2 mM e 0.4 mM, FeSO_4 1.0 mM e 2.0 mM, ZnSO_4 0.02 mM, MnSO_4 0.8 mM e CoSO_4 0.2 mM observa-se que geralmente a linhagem nocaute apresenta $\text{OD}_{600\text{nm}}$ ligeiramente maior que as apresentadas pela linhagem controle (Figura 4.5). Apesar das diferenças serem pequenas, este padrão pode indicar que a adição de metais melhora o crescimento da linhagem, sugerindo alterações relacionadas ao metabolismo envolvendo esses metais de transição ou então, um possível efeito dos fatores que incrementam a $\text{OD}_{600\text{nm}}$ nas etapas finais do crescimento: liberação de vesículas de gás e produção de bacteriorodopsina.

No anexo 7.1 se pode consultar as curvas de crescimento nas que se avalia as restantes concentrações dos metais de transição.

4.3.2.2 Estresse oxidativo

Além de testar metais de transição, examinamos as possíveis mudanças fenotípicas ocasionadas por dois agentes oxidantes em *H. salinarum* NRC-1: o H_2O_2 que facilita a formação de radicais hidroxila ($\cdot\text{OH}$) e o PQT que, mediante ciclos redox, promove a formação de radicais superóxido (O_2^-). As respostas transcricionais a esses agentes foi mapeada e sabe-se que frente a $\cdot\text{OH}$, há um aumento dos níveis dos transcritos associados com as peroxidases; já frente ao O_2^- , além das peroxidases, há um aumento dos níveis de transcritos associados com superóxidos desmutases. Sistemas alternativos ou secundários também são ativados para sequestrar espécies reativas do oxigênio (EROs) como antioxidantes (carotenóides), rodopsinas (bacteriorodopsina, halorodopsina, sensores da rodopsina I e II), além da supressão da biogê-

Figura 4.5: Curvas de crescimento onde foram detectadas ligeiras diferenças de OD_{600nm} entre as linhagens controle (Aura3-vermelho) e nocaute (Δ VNG_RS05855-preto) sob a influência de CuSO₄, FeSO₄, ZnSO₄, MnSO₄ e CoSO₄. As linhas verticais representam o desvio padrão.

nese das vesículas de gás (Baliga et al., 2002; Kaur et al., 2010).

Frente ao PQT as linhagens não apresentam diferenças nas medições de OD_{600nm} durante o crescimento. Observa-se a diminuição drástica do crescimento ainda nas menores concentrações avaliadas (Figura 4.6). Um resultado similar foi obtido por Kaur e col ao testar as mesmas concentrações de PQT na linhagem selvagem *H. salinarum* NRC-1 (Kaur et al., 2010). Em H₂O₂ 5mM identificamos uma diferença muito pequena de OD_{600nm} a partir de 44 horas de crescimento, com a linhagem nocaute apresentando as menores OD_{600nm} (Figura 4.6). O

efeito é pequeno e poderia indicar uma maior sensibilidade ao agente estressante na linhagem nocaute.

Figura 4.6: Curvas de crescimento realizada nas linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) frente a diferentes concentrações do agente oxidante 1,1'-dimetil-4,4'-bipiridina-dicloreto (PQT) e peróxido de hidrógeno (H_2O_2). As linhas verticais representam o desvio padrão dos dados.

4.3.2.3 Arsênio

Avaliamos também o efeito fenotípico ocasionado pelo arsenito de sódio ($NaAsO_2$) nas linhagens. Este metalóide se liga a grupos sulfidril das proteínas e inibe a oxidação do piruvato, afetando o ciclo dos ácidos tricarboxílicos, reduzindo a fosforilação oxidativa e afetando a gliconeogênese (Carol M. Schiller, Bruce A. Fowler e Woods, 1977; Kulshrestha, 2014). Em *H. salinarum* NRC-1 foram descritos genes homólogos aos operons das bactérias que conferem resistência a arsênio (Gejiao Wang, Sean P. Kennedy, Sabeena Fasiludeen, Christopher Rensing e DasSarma, 2006). Em bactérias, a expressão diferencial de seus membros é associado com mecanismos de regulação pós-transcricionais (Owolabi e Rosen, 1990).

Nossos resultados mostram diferenças consistentes na concentração de 0.5mM, onde a linhagem nocaute apresenta menor OD_{600nm} do que a linhagem controle ao longo de toda a curva de crescimento. Nas outras concentrações testadas, as curvas apresentam perfis semelhantes, porém a dispersão dos dados foi muito grande (Figura 4.7).

Figura 4.7: Curvas de crescimento das linhagens controle (Δura3 -vermelho) e nocaute ($\Delta\text{VNG_RS05855}$ -preto) em diferentes concentrações do metalóide arsenito de sódio (NaAsO_2). As linhas verticais representam o desvio padrão dos dados.

4.3.2.4 Concentração de sal

Baseados nos estudos feitos por Beer e col, onde foram examinadas variações na composição do meio CM padrão do crescimento de *H. salinarum* NRC-1 (Beer, Wurtmann e Baliga, 2014), testamos duas composições distintas: 86% de sais em relação ao meio CM (meio A), e 72% (meio B). A diminuição de sais no meio de crescimento de *H. salinarum* NRC-1 promove mudanças nas taxas de crescimento, na morfologia da célula, conteúdo de vesículas de gás e regulação de genes vinculados ao metabolismo anaeróbico (Beer, Wurtmann e Baliga, 2014; Vauclare et al., 2015).

No meio A, observamos que a partir de 50 horas de crescimento a linhagem nocaute apresenta menor $\text{OD}_{600\text{nm}}$ do que a linhagem controle (Figura 4.8). Isto pode indicar uma maior sensibilidade a variações na concentração de sal na transição entre as fases exponencial e estacionária do crescimento, onde o metabolismo de *H. salinarum* NRC-1 pode mudar de aeróbico a anaeróbico (Beer, Wurtmann e Baliga, 2014). Também a diminuição da $\text{OD}_{600\text{nm}}$ poderia ser o reflexo de uma diferença na produção de vesículas de gás entre as linhagens. No meio B é possível distinguir que com a diminuição de sais no meio cai a taxa de crescimento de ambas linhagens com um perfil similar ao que se distingue no meio A. Porém, só se detectam

pequenas diferenças entre as linhagens a partir das 91 horas (Figura 4.8).

Figura 4.8: Curvas de crescimento das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) em meios com uma concentração inferior de sal ao meio CM. Meio A mantém o 86% de sais e o meio B o 72% dos sais respeito às contidas no meio CM de referência para a cultivo da *H. salinarum* NRC-1. As linhas verticais representam o desvio padrão dos dados.

A avaliação dos diferentes estresses durante a curva de crescimento permitiu identificar condições em que a linhagem nocaute apresentou OD_{600nm} diferente da linhagem controle, indicando que a deleção do VNG_RS05855 pode influenciar de forma sutil na resposta a diferentes estresses. Para quatro metais de transição (Zn^{2+} , Cu^{2+} , Fe^{2+} , Mn^{2+}), observamos que a linhagem nocaute apresentou OD_{600nm} maior do que o controle nas fases finais de crescimento; o oposto foi observado em H_2O_2 , arsênico e na modificação do meio CM no que se manteve o 86% dos sais (meio A). De forma geral, as diferenças observadas foram pequenas, exceto para arsênico 0.5mM. Durante a aquisição dos dados, notamos que a dispersão foi grande entre as diferentes réplicas o que nos levou a obter uma grande quantidade de dados. A grande dispersão dos dados tem sido descrita na literatura em outras linhagens derivadas de *H. salinarum* NRC-1 (Tonner et al., 2017), sugerindo que a variação poderia ter componentes experimentais e fisiológicos. Para verificar se estes efeitos observados podem ser de fato fisiologicamente relevantes, optamos por analisar a diferença nas medidas de absorbância entre as linhagens, na busca de padrões confiáveis.

4.3.3 As diferenças nas medições de absorbância se concentram nas fases finais do crescimento

Com base no elevado número de curvas de crescimento realizadas das linhagens con-

trole e nocaute e a manutenção dos horários nas que foram efetuadas as medições das absorbâncias, decidimos implementar uma análise complementar para identificar fenótipos associados à linhagem nocaute. Realizamos a subtração das medições de absorbâncias de todas as réplicas das linhagens nocaute (Δ VNG_RS05855) e das da linhagem controle (Δ ura3) nos mesmos tempos do crescimento e denominamos o resultado como diferença de OD_{600nm}. Com esta abordagem determinamos as diferenças de OD_{600nm} nas condições de crescimento padrão e sob estresses, com o intuito de ressaltar as diferenças entre as linhagens que poderiam passar despercebidas nos gráficos das curvas de crescimento.

Nas condições de crescimento padrão para a estimativa da diferença de OD_{600nm} juntamos as medições de absorbâncias obtidas do crescimento em placa (200 μ L) e em frasco (40 mL) das distintas linhagens até às 100 horas de crescimento, por ser o menor tempo na que se cultivaram as linhagens em frascos. Observamos que as médias das medições de absorbância nas etapas finais do crescimento tem um perfil superior na linhagem nocaute (Figura 4.9 A), no entanto este resultado não reproduz o obtido durante o início da caracterização fenotípica (Figura 4.4), onde esta linhagem não mostrou incrementos no perfil de absorbâncias em relação a linhagem controle. Uma possível explicação a esta variabilidade é o aumento na dispersão dos dados a medida que são consideradas um maior número de réplicas, sendo um comportamento observado em outras linhagens de *H. salinarum* NRC-1 (Tonner et al., 2017).

Nas fases iniciais das curvas de crescimento, observa-se que os valores de diferença de OD_{600nm} são ligeiramente negativos ou próximos a zero, indicando que os valores de OD_{600nm} da linhagem controle são levemente superiores ou similares aos da linhagem nocaute. Isso pode apontar a uma ligeira desaceleração do crescimento da linhagem nocaute nas etapas de iniciais do crescimento que desaparece uma vez que se avança no crescimento. As maiores variações de diferenças de OD_{600nm} aparecem nas etapas finais do crescimento, sendo os valores superiores a zero indicativos do incremento da OD_{600nm} na linhagem nocaute em relação ao controle (Figura 4.9 B). Assim, na ausência do gene VNG_RS05855 os aumentos na diferença de OD_{600nm} indicam a fenótipos distintos entre as linhagens. Porém a dispersão dos dados também é maior, o que pode refletir em diferenças não no crescimento, e sim na possibilidade de uma diferença na produção de vesículas de gás ou compostos fotossensíveis.

Figura 4.9: Conjunto total de medições de absorbância a 600 nm das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) durante o crescimento nas condições referência obtidas de culturas crescidas em 200 μ L e 40 mL de meio CM. A) Variações da OD_{600nm} durante o tempo (eg. curva de crescimento). Os pontos representam os valores de absorbâncias, as linhas contínuas as médias desses valores e os sombreados os desvios padrões das medições. B) Representação das medições de diferença de absorbâncias (diferença OD_{600nm}), estimada como a subtração, nos mesmos tempos, dos valores de absorbância das linhagens nocaute (Δ VNG_RS05855) menos a controle (Δ ura3-vermelho).

Para as condições de crescimento sob estresses testadas colocamos todas as medições de absorbâncias das réplicas das linhagens obtidas do total de estresses (Figura 4.10 A). Observamos não só a alta dispersão dos dados, o que pode indicar a falta de reprodutibilidade, como também o desaparecimento dos incrementos no perfil de absorbância da linhagem nocaute nas etapas finais do crescimento; como foi observado durante o crescimento nas condições padrões (Figura 4.9 A). Isso pode indicar a susceptibilidade da linhagem nocaute perante os estresses e a possível interconexão entre eles e o gene VNG_RS05855 com os fatores que promovem os incrementos nas absorbâncias nas etapas finais do crescimento em *H. salinarum* NCR-1.

Ao avaliar o comportamento das diferenças de OD_{600nm} observamos que são parcialmente reprodutíveis a partir das 60 horas de crescimento frente aos metais de transição, especificamente nas três menores concentrações de Cu²⁺, Zn²⁺, Fe²⁺ (Figura 4.10 B até D), em todas as concentrações testadas de Mn²⁺ (Figura 4.10 F) e valores dispersos com Co²⁺ (Figura 4.10 G). Neles a diferença de OD_{600nm} se caracteriza por serem valores positivos. A susceptibilidade da linhagem nocaute ao arsenito de sódio (NaAsO₂) é sugerida pelo os valores negativos da diferença de OD_{600nm} (Figura 4.10H), indicando a possibilidade de VNG_RS05855 estar relacionado com a resposta adaptativa a este metalóide. No estresse oxidativo induzido por PQT a

diferença de OD_{600nm} é próximo a zero, sugerindo que não existem diferenças fenotípicas entre as linhagens frente a este estresse (Figura 4.10 E). Com o H₂O₂ o comportamento é distinto, na menor concentração (5 mM) a diferença na OD_{600nm} tiveram valores negativos e para a concentração maior (25 mM) as medições de diferença de OD_{600nm} foram positivas (Figura 4.10 I). Ao avaliar como as mudanças nas concentrações de sais do meio CM valores negativos após das 60 horas de crescimento, já em tempos menores os valores são próximos de zero (Figura 4.10 J).

Figura 4.10: Conjunto total de medições de absorbância a 600 nm das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) durante o crescimento nos diferentes estresses obtidas de culturas crescidas em 200 μ L e 40 mL de meio. A) Variações da OD_{600nm} durante o tempo (eg. curva de crescimento). Os pontos representam os valores de absorbâncias de todas as réplicas das linhagens controle e nocaute submetidas ao conjunto de estresses, as linhas contínuas as médias desses valores e os sombreados o desvio padrão dos dados. B até J) Representação das medições de diferenças de absorbâncias estimada como a subtração, nos mesmos tempos, dos valores de absorbância das linhagens nocaute (Δ VNG_RS05855) menos a controle (Δ ura3) nas diferentes concentrações dos distintos estresses. B) CuSO_4 (0.2, 0.4, 0.8, 1.0, 1.5, 2.0 mM); C) ZnSO_4 (0.02, 0.05, 0.1, 0.15, 0.20, 0.25 mM); D) FeSO_4 (1.0, 2.0, 4.0, 5.0, 6.0, 8.0 mM); E) PQT (0.125, 0.250, 0.500, 0.750, 1.000, 1.125 mM); F) CoSO_4 (0.2, 0.5 mM); G) MnSO_4 (0.8, 1.5 mM); H) NaAsO_2 (0.1, 0.2, 0.5 mM); I) H_2O_2 (5, 25 mM); J) Meio A (NaCl-216 g/L, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ -16.8 g/L, Na-Citrate-2.7 g/L, KCl-1.8 g/L, e peptone 10 g/L) e Meio B (NaCl-180 g/L, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ -14 g/L, Na-Citrate-2.3 g/L, KCl-1.5 g/L, e peptone 10 g/L)

O conjunto dos resultados das curvas de crescimento nos permite propor que a ausência do gene VNG_RS05855 apresenta uma influência sutil nas medidas de absorbância. As sutis diferenças fenotípicas com a linhagem controle sugere que o seu produto proteico esteja relacionado a fatores que incrementam a absorbância nas etapas finais do crescimento e/ou ao processamento de transcritos associados com metais, os de respostas a arsênico e os vinculados ao estresse oxidativo.

4.3.4 As diferenças de absorbância nas etapas finais do crescimento não está associada com o acúmulo diferencial de bacteriorodopsina entre as linhagens

Os incrementos da absorbância em *H. salinarum* NRC-1 nas fases finais do crescimento têm sua origem em dois fatores principais: (1) o aumento na produção de compostos fotossensíveis como a bacteriorodopsina e (2) incremento na liberação ao meio extracelular das vesículas de gás, ambos fatores, capazes de contribuir com a absorção a 600 nm. O próprio crescimento celular poderia ser considerado um terceiro fator, porém já foi mostrado que o incremento na densidade celular não são proporcionais aos aumentos da OD_{600nm} nas fases finais do crescimento (Facciotti et al., 2010; Yao e Facciotti, 2011)(Figura 4.4).

Para estudar a possível influência do produto de VNG_RS05855 na acúmulo diferencial de bacteriorodopsina nas linhagens controle e nocaute realizamos medições das absorbâncias em diferentes comprimentos de onda. A bacteriorodopsina é capaz de absorver a luz nos comprimentos de onda 532 e 600 nm contribuindo com as medições absorbância durante o crescimento. No entanto, em outros comprimentos de onda como 700 e 785 nm, esta molécula não consegue absorver luz (Facciotti et al., 2010).

Durante o crescimento das linhagens obtivemos os maiores valores de densidade óptica nos comprimentos de onda 532 e 600 nm (OD_{532nm} e OD_{600nm}), onde a bacteriorodopsina é capaz de absorver luz (Figura 4.11 A, B). No entanto, pela alta dispersão das medições, é difícil identificar diferenças entre elas. A dispersão da linhagem controle é sempre maior que na linhagem nocaute o que nos pode sugerir que a produção de compostos fotossensíveis nesta linhagem tenha uma maior variabilidade. Nos comprimentos de onda de 700 e 785 nm (OD_{700nm} e OD_{785nm}) as curvas de crescimento se caracterizam por apresentar valores menores

da densidade óptica, sem diferenças relevantes do ponto de vista estatístico e um perfil similar ao encontrado com os restantes comprimentos de onda (Figura 4.11 C, D). Este último resultado nos sugere que não há diferenças entre as linhagens no acúmulo da bacteriorodopsina.

Figura 4.11: Medição da absorbância a diferentes comprimentos de onda das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto). A) 532 nm; B) 600 nm; C) 700 nm; D) 785 nm. As linhas verticais representam o desvio padrão dos dados.

4.3.5 Ausência do gene VNG_RS05855 influencia as transições metabólicas diante de mudanças na disponibilidade de nutrientes

A característica da bacteriorodopsina de absorver a energia contida nos fótons da luz é chave em *H. salinarum* NRC-1 no processo de obtenção de energia química em condições limitantes de oxigênio (DasSarma et al., 2012). Nestas condições, outros processos para a formação de ATP podem ser ativados como a respiração anaeróbica ou a fermentação (Ng et al., 2000). No entanto, só a fotofosforilação e respiração podem promover a acidificação do meio extracelular (Oesterhelt e Stoeckenius, 1973; Oesterhelt e Krippahl, 1973), que também é favorecido pelo gradiente eletroquímico de prótons proposto por Mitchell na sua teoria quimiosmótica (Mitchell, 1961).

Com base nestas características do metabolismo de *H. salinarum* NRC-1, verificamos se a ausência do gene VNG_RS05855 afeta as transições fisiológicas entre os mecanismos de obtenção de energia. Para isso realizamos medições do pH do meio extracelular e determinamos

se as suas possíveis variações, entre as linhagens, seriam influenciadas pela luz.

As primeiras medições de pH foram feitas em diferentes momentos do crescimento, tanto da linhagem controle como a nocaute. Realizamos também medições do pH do sobrenadante de cada linhagem após das culturas serem centrifugadas. Os resultados mostram que o pH é mais ácido na cultura que no sobrenadante (Figura 4.12), confirmando a presença nas membranas celulares de *H. salinarum* NRC-1 de componentes que influenciam a acidez de meio extracelular (Ng et al., 2000; Talaue et al., 2016; WAGNER, HARTMANN e OESTERHELT, 1978). Entre as linhagens se distinguem diferenças estatisticamente significativas às 65 horas de crescimento, sendo o meio extracelular da linhagem nocaute mais básico, indicando que neste momento do crescimento a troca de hidrogênio com o meio possa ter alguma relação com VNG_RS05855.

Outro aspecto interessante é o aumento do pH do meio extracelular no transcurso do crescimento das linhagens. A alcalinização do meio extracelular indica que o interior da célula, à medida que aumenta o tempo de cultura, se acidifica (WAGNER, HARTMANN e OESTERHELT, 1978; Slonczewski et al., 2009). Precisamente, o prótons que as bacteriorodopsinas e/ou a fosforilação oxidativa colocam para fora da célula são internalizados pelas ATP sintases na produção do ATP, visto que a medida que avança o crescimento, o processo de produção de energia se torna crítico na manutenção da população celular. Processos similares de alcalinização do meio de cultivo a medida que avança o crescimento é um comportamento previamente relatado para *H. salinarum* isolados de lagos salinos no Iran (Hassanshahian e Mohamadian, 2011) e para *E. coli* (Sánchez-Clemente et al., 2018).

O próximo passo foi estudar o efeito sobre o pH de diluições das culturas das linhagens em diferentes etapas do crescimento. Alíquotas das linhagens foram diluídas em solução basal de sal e em meio de cultura CM, a diferença entre elas é a disponibilidade de nutrientes a serem usadas na obtenção de energia pelas células. As diluições em solução basal promovem a obtenção de energia a partir da luz, já que nas diluições não se disponibilizam nutrientes que permitam a produção de ATP. Nesses casos, as principais variações do pH deveriam estar relacionadas com a bacteriorodopsina. A diluições em meio CM, tem como objetivo, avaliar as variações de pH pelo conjunto de reações da cadeia respiratória, que aproveitam produtos gera-

Figura 4.12: Medições de pH das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) durante o crescimento A) na cultura celular; B) no sobrenadante. As linhas verticais representam o desvio padrão dos dados.

dos durante a degradação dos nutrientes do meio, para produzir ATP (Ng et al., 2000; Gonzalez et al., 2009; Talaue et al., 2016). Em cada diluição o efeito da luz foi avaliado mantendo as alíquotas na luz ou no escuro.

Nas alíquotas em que se diminui a disponibilidade de nutrientes mediante diluições em solução basal observamos diferenças entre os pH das linhagens às 41 e 95 horas de crescimento na luz; no escuro essas diferenças desaparecem (Figura 4.13 A e B, respectivamente). No meio com restrição transitória de nutrientes, as variações de pH entre as linhagens parece ser influenciada pela geração do gradiente de prótons promovido pela fototrofia. No entanto, o efeito da ausência de VNG_RS05855 é diferente, deixando o meio extracelular da linhagem nocaute mais básico às 41 horas que a linhagem controle e às 95 horas mais ácido; sugerindo que nestas condições a sua função depende da fase do crescimento.

Ao avaliar as mudanças no pH nas linhagens após as diluições nas que se disponibilizam todos os nutrientes, observamos que não existem diferenças significativas nas medições do pH na luz (Figura 4.14 A). Orland e col propõem que durante a fototrofia, ao contrário do que é normalmente assumido, também se disponibiliza energia metabólica mediante a fermentação de arginina (Baliga et al., 2002; Gonzalez et al., 2009). Neste sentido, sugerimos que a presença vários mecanismos de obtenção de energia equilibre as demandas energéticas das células e mantenha, entre as linhagens, indistinguíveis as possíveis mudanças do pH que poderiam ser originadas pela fototrofia.

Ao disponibilizar todos os nutrientes nas diluições das linhagens as medições dos pH

Figura 4.13: Medições de pH das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) diluídas em solução basal de sal mantidas na A) luz; B) no escuro em distintos pontos da curva de crescimento. As linhas verticais representam o desvio padrão dos dados e o asterisco o nível de significância estatística segundo o teste *t-student* ((*) $p < 0.05$)

Figura 4.14: Medições de pH das diluições em meio CM das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) mantidas na A) luz; B) no escuro. As linhas verticais representam o desvio padrão dos dados e o asterisco o nível de significância estatística segundo o teste *t-student*((*) $p < 0.05$; (**) $p < 0.01$; (***) $p < 0.001$)

são diferentes significativamente no escuro às 17, 65 e 95 horas do crescimento (Figura 4.14 B). Sem luz não ocorre fototrofia, e o processo que pode alterar o pH é a cadeia de transporte de elétrons da respiração. A diferença de pH, deixando o meio extracelular da linhagem nocaute mais ácido que a linhagem controle, sugere que o efeito da ausência de VNG_RS05855 seja sobre os transcritos associados ao transporte de elétrons (WAGNER, HARTMANN e OESTERHELT, 1978; Talaue et al., 2016).

4.3.6 Os principais biomarcadores *H. salinarum* NRC-1 não são afetados pela ausência do gene VNG_RS05855

A integração de dados genômicos e do proteoma em *H. salinarum* NRC-1 revelou que

a regulação da fototrofia não ocorre unicamente pela mediação da montagem da bacteriorodopsina, outras vias metabólicas também estão co-reguladas como a síntese de isoprenóides e outros carotenóides (Baliga et al., 2002). Os principais derivados de carotenóides presentes nas haloarchaeas são pigmentos complexo retinal-proteína, bacteriorodopsina, e a bacterioruberina. Os carotenóides atuam como potentes antioxidantes, reguladores das mudanças osmóticas das células e são utilizados como biomarcadores para a identificação de linhagens nas haloarchaeas (Marshall et al., 2006). Esta última aplicação tem sido realizada como emprego da técnica espectroscópica Raman (Marshall et al., 2006; Fendrihan, Musso e Stan-Lotter, 2009) que foi utilizada, neste estudo, para identificar se há algum comportamento diferencial entre as bioassinaturas das linhagens controle e nocaute.

Os espectros Raman dos carotenóides são dominados por picos em 1003 cm^{-1} , 1151 cm^{-1} e 1507 cm^{-1} e correspondem, respectivamente, com os deslocamentos dos grupos metilas ligados no mesmo plano a cadeias de polienos (C-CH₃), ligações simples carbono-carbono (C-C) e as duplas ligações conjugadas (C=C) (James). Nas haloarchaeas existem uma variedade de compostos carotenóides e seus espectros são semelhantes, porém não idênticos. A correlação entre a posição exata do pico de distensão das ligações duplas conjugados (C=C) e o comprimento de conjugação da cadeia principal de polieno permite distingui-los (Ames et al., 1992).

Após excitar as amostras de diferentes etapas do crescimento das linhagens controle e nocaute com 532 nm e 785 nm para aumentar a intensidade do sinal Raman e diminuir o efeito da fluorescência de fundo, respectivamente, focamos em estudar a fração espectral entre 600 cm^{-1} e 1600 cm^{-1} .

Observamos que a sobreposição da média dos espectros das linhagens são dominados pelas características dos picos dos carotenóides, embora não conseguimos identificar neles diferenças significativas em nenhuma das etapas do crescimento avaliadas (Figura 4.15 A até D e Figura 4.16 A até D).

A partir do conjunto de réplicas biológicas e experimentais nos diferentes tempos de crescimento, realizamos a subtração dos espectros das linhagens nocaute e controle, num esforço para melhorar a identificação das diferenças. As diferenças em ambos são quase sempre

zero, refletindo a semelhança entre os espectros das linhagens (Figura 4.15 E até H e 4.16 E até H).

Contudo, é interessante observar o comportamento na fração espectral 1250 cm^{-1} - 1450 cm^{-1} . Esta região tem sido associada nos espectros Raman aos ácidos graxos em membranas de organismos eucariotos e bactérias (Fendrihan, Musso e Stan-Lotter, 2009), porém arqueias não possuem ácidos graxos nas suas membranas, formadas por derivados de isoprenos, o que caracteriza as intensidades Raman nestas regiões como geralmente fracas (Pande, Lanyi e Callender, 1989; Ames et al., 1992; Marshall et al., 2006; Fendrihan, Musso e Stan-Lotter, 2009). Embora este resultado seja uma surpresa, conseguimos atribuir componentes moleculares aos picos mais proeminentes nesta região. Assim, o pico 1200 cm^{-1} se associa com a região dos carotenóides, o pico 1285 cm^{-1} se associa com grupos amida III (CNH) das proteínas, com simples enlaces carbono-carbono (C-C) e carbono-carbono-hidrogênio (C-C-H); e o pico $\sim 1448\text{ cm}^{-1}$ com grupo molecular CH_2 que pode formar parte de lípidos, proteínas e carboidratos, fundamentalmente dos constituintes das membranas (Ames et al., 1992). Mesmo com esta associação, o comportamento dos espectros nesta fração continua sendo um mistério.

Figura 4.15: Espectroscopia Raman das linhagens controle (Δ aura3-vermelho) e nocaute (Δ VNG_RS05855-preto) obtidos após de excitação das amostras a 532 nm. A-D) Superposição das médias dos espectros Raman obtidos do total de réplicas das linhagens Δ aura3 e Δ VNG_RS05855 nas 17, 41, 65 e 95 horas, respectivamente. O sombreado cinza representa o desvio padrão de ter trabalhado com três réplicas biológicas. E-H) Subtração das médias dos espectros do total de réplicas da linhagem Δ VNG_RS05855 vs a linhagem Δ aura3 às 17, 41, 65 e 95 horas, respectivamente.

As interações entre carotenóides diferentes dentro da biomassa celular afeta as cadeias principais de polienos e pode causar deslocamentos significativos dos picos (cm^{-1}) devido a mudanças na deslocalização eletrônica (Oliveira et al., 2010; Jehlička, Edwards e Oren, 2014). Talvez esta possa ser uma explicação para o comportamento na fração espectral 1250 cm^{-1} - 1450 cm^{-1} . No entanto, nós assumimos que os picos característicos dos biomarcadores de halorarchaeas deveriam aparecer inalteráveis quando eles tivessem uma produção diferencial em amostras complexas. Os biomarcadores mais abundantes, seriam menos afetados pela deslocalização eletrônica promovida por espécies menos abundantes.

Desta maneira, identificamos nos espectros Raman das linhagens controle e nocaute, obtidos com excitação a 532 nm e 785 nm, as características dos picos Raman reportadas na literatura dos principais biomarcadores das haloarchaeas. Estudamos as características de carotenóides, a bacteriorodopsina (Jehlička e Oren, 2013), a halorodopsina (Ames et al., 1992) e o sensor da II da halorodopsina (Person et al., 2005). Halorodopsina é também uma bomba acionada por luz para transportar ânions (Cl^- , fundamentalmente) e o SRII é receptor da luz que ativa a fototaxia para que a célula procure o escuro quando a respiração e os nutrientes são abundantes (Person et al., 2005).

As intensidades dos picos Raman que caracterizam os biomarcadores estudados não mostraram diferenças significativas nos espectros obtidos nas linhagens controle e nocaute, tanto na excitação em 532 nm como a 785 nm. Mostramos os resultados das médias das réplicas biológicas para cada linhagens às 41 e 95 horas de crescimento (Figura 4.17).

Distinguimos que na linhagem nocaute a maioria as intensidades Raman dos picos que caracterizam a bacteriorodopsina e halorodopsina tem uma tendência a serem superiores aos obtidos na linhagem controle (Figura 4.17 C até F).

Figura 4.17: Intensidades Raman dos biomarcadores de *H. salinarum* NRC-1 nas linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) às 41 e 95 horas de crescimento. A-B) Carotenóides (Jehlička e Oren, 2013) obtidos com excitação a 532 e 785 nm, respectivamente; C-D) Bacteriorodopsina (Jehlička e Oren, 2013) obtidos com excitação a 532 e 785 nm, respectivamente; E) Halorodopsina excitada a 531 nm (Ames et al., 1992) em nossos dados excitação a 532 nm; F) Halorodopsina excitada a 752 nm (Ames et al., 1992) em nossos dados excitação a 785 nm; G-H) Sensor II da rodopsina-SRII (Person et al., 2005) obtidos com excitação a 532 e 785 nm, respectivamente. As intensidades Raman se correspondem com a média do total de réplicas empregadas neste ensaio e as linhas verticais o seu desvio padrão.

4.3.7 A produção de vesículas de gás entre as linhagens nocaute e controle não apresentam diferenças relevantes

Com a estimativa das diferenças de OD_{600nm} identificamos o incremento da absorbância nas fases finais do crescimento na linhagem nocaute em relação a linhagem controle. Um possível fator que influencia esta diferença é a produção de compostos fotossensíveis. Nós realizamos vários ensaios para testar esta hipótese, porém nossos resultados não foram categóricos para assumir esta relação. Os incrementos das absorbâncias nas etapas finais do crescimento em *H. salinarum* NRC-1, também são relacionados à liberação ao meio extracelular de vesículas de gás (Facciotti et al., 2010; Yao e Facciotti, 2011).

Nós não conseguimos identificar a presença das vesículas de gás no meio extracelular nas linhagens controle e nocaute por microscopia ótica, como foi realizado por Facciotti e col ao estudar a linhagem selvagem de *H. salinarum* (Facciotti et al., 2010). No entanto, estudamos as diferenças na produção interna de vesículas de gás nas linhagens controle e nocaute realizando ensaios de microscopia eletrônica de transmissão.

As imagens do espaço intracelular revelam que nas células das linhagens controle e nocaute a produção destas vesículas de gás aumenta a durante o crescimento (Figura 4.18). Contudo, mediante inspeção visual, não conseguimos distinguir diferenças entre a quantidade destas estruturas supramoleculares entre as linhagens nas distintas etapas do crescimento. Em 17 horas, é possível observar uma diferença sutil entre o número de vesículas nas células da linhagem, com um maior número na linhagem nocaute em relação ao controle. No entanto, quantificações mais precisas deverão ser realizadas utilizando mais imagens e softwares para quantificar estas estruturas.

Andrew e col 2011 mostram que as variações de absorbâncias promovidas pelas vesículas de gás livres no meio extracelular são proporcionais ao que poderia ser esperado pela lise celular nas fases tardias do crescimento (Yao e Facciotti, 2011). Em resultados prévios onde avaliamos as variações de densidade celular mostramos que não existem diferenças no número de células entre as linhagens durante o crescimento (Figura 4.4) sugerindo que se o número de células permanece inalterado, possivelmente, porque ambas linhagens compartilham as características de integridade celular. Se consideramos que o número de vesículas de gás internos

Figura 4.18: Microfotografias das linhagens Δ ura3 (controle) e Δ VNG_RS05855 (nocaute) às 17, 41, 65 e 95 horas do crescimento. Marcadas às 17hrs as estruturas das vesículas de gás em ambas linhagens que são observadas nos posteriores tempos.

entre as linhagens não varia, podemos sugerir que as mudanças na OD_{600nm} relativas nas etapas finais do crescimento entre as linhagens não se justificam pela liberação das vesículas de gás ao meio extracelular.

4.3.8 Diante das concentrações semi-letais de agentes estressantes, as linhagens têm o mesmo nível de sobrevivência

Os ensaios de sobrevivência consistem em submeter as células a um estresse drástico e depois plaquear diferentes diluições em placas Petri com meio sólido. Após o estresse somente as células viáveis originam colônias.

Nós utilizamos este experimento para avaliar o efeito da ausência do gene VNG_RS05855 na viabilidade celular após das células serem submetidas a concentrações semi-letais de agentes estressantes.

Os ensaios de sobrevivência foram realizadas nas linhagens controle e nocaute na fase exponencial do crescimento ($OD_{600nm} \sim 0.5$) utilizando concentrações semi-letais agentes estressantes Cu^{2+} , AsO_2^- e PQT por serem os estresses onde foi possível distinguir características

distintivas do comportamento das diferenças de OD_{600nm} nas linhagens.

Os resultados não permitem distinguir diferenças significativas na sobrevivência das linhagens após do tratamento com nenhum dos agentes estressantes (Figura 4.19 até 4.21). Estes resultados sugerem que VNG_RS05855 não influencia as respostas adaptativas a estes estresses e que garantam a viabilidade celular diante do condicionamento a concentrações semi-letais de Cu²⁺, AsO₂⁻ e PQT.

Figura 4.19: Ensaios de sobrevivência das linhagens Δura3 (controle) e ΔVNG_RS05855 (no-caute) submetido a condição semi-lethal de 1.25 mM de CuSO₄. A) Fotografias das placas de uma réplica biológica nos diferentes tempos de exposição ao agente estressante e as distintas diluições das culturas após este tempo. B) Média da contagem das unidades formadoras de colônias (UFC- assumindo com o número de pixels da colônia na placa) nas diluições das culturas nos tempos de exposição ao agente estressante. As barras pretas são desvio padrão das medições com base no conjunto total de réplicas. Tanto em A) como B) é apresentado o comportamento das linhagens sem estresses.

Figura 4.20: Ensaios de sobrevivência das linhagens *Δura3* (controle) e *ΔVNG_RS05855* (no-caute) submetido a condição semi-letal de 0.3mM de NaAsO₂. A) Fotografias das placas de uma réplica biológica nos diferentes tempos de exposição ao agente estressante e as distintas diluições das culturas após este tempo. B) Média da contagem das unidades formadoras de colônias (UFC- assumindo com o número de pixels da colônia na placa) nas diluições das culturas nos tempos de exposição ao agente estressante. As barras pretas são desvio padrão das medições com base no conjunto total de réplicas. Tanto em A) como B) é apresentado o comportamento das linhagens sem estresses.

Figura 4.21: Ensaios de sobrevivência das linhagens *Δura3* (controle) e *ΔVNG_RS05855* (no-caute) submetido a condição semi-letal de 4.0mM de PQT. A) Fotografias das placas de uma réplica biológica nos diferentes tempos de exposição ao agente estressante e as distintas diluições das culturas após este tempo. B) Média da contagem das unidades formadoras de colônias (UFC- assumindo com o número de pixels da colônia na placa) nas diluições das culturas nos tempos de exposição ao agente estressante. As barras pretas são desvio padrão das medições com base no conjunto total de réplicas. Tanto em A) como B) é apresentado o comportamento das linhagens sem estresses.

Com o conjunto dos resultados dos ensaios fenotípicos podemos sugerir que a ausência do gene VNG_RS05855 afeta a resposta de *H. salinarum* NRC-1 frente a determinados estímulos ambientais: metais de transição (Zn^{2+} , Cu^{2+} , Fe^{2+} , Mn^{2+} , Co^{2+}), um metalóide (NaAsO_2), um agentes oxidantes (H_2O_2) e duas variações de sais do meio CM. As diferenças em relação à linhagem controle são sutis e a sua ausência não apresenta um efeito dramático nos fatores que promovem o incremento da $\text{OD}_{600\text{nm}}$ nas etapas finais do crescimento em *H. salinarum* NRC-1. No entanto, propomos que apesar de não ser essencial no desenvolvimento deste organismo, a atividade do produto proteico do gene VNG_RS05855 possa estar relacionada a transições fisiológicas do metabolismo energético, talvez como um mecanismo secundário de apoio ao crescimento.

4.4 Transcritoma das linhagens controle e nocaute

A comparação de transcritomas de linhagens diferentes de uma mesma espécie tem apoiado a caracterização das diferenças fenotípicas, por intermédio da identificação de alterações na expressão de genes relacionados com vias metabólicas e/ou mecanismos de resposta a estresses (Gerstein et al., 2014; Wurtmann et al., 2014; Gomes-Filho et al., 2015; Ten-Caten et al., 2018).

Com base neste pressuposto realizamos o sequenciamento dos transcritomas das linhagens controle e nocaute em $\text{OD}_{600\text{nm}} \sim 0.5$, às 41 horas de crescimento, com diferentes modos de preparo das bibliotecas e distintas amostras de RNA. Preparamos as bibliotecas com os protocolos TruSeq Small e TruSeq mRNA Stranded. Das bibliotecas preparadas com o protocolo TruSeq Small, exploramos a potencialidade de conservar melhor a informação da especificidade das fitas, conforme prévia experiência em nosso laboratório; com o TruSeq mRNA Stranded, exploramos todos os RNAs visto que não houve seleção do RNA por tamanho. Os protocolos foram utilizados em diferentes tipos de amostras: amostras de RNA fragmentado, RNA sem fragmentar e RNA enriquecidas com RNA primário. As duas últimas amostras foram usadas no estudo da expressão gênica diferencial entre as linhagens e na predição de possíveis sítios de processamento, respectivamente, as amostras de RNA fragmentado foram empregadas para

estudar o provável perfil de degradação dos RNA nas linhagens. Mais detalhes e justificativas das aplicações destes procedimentos serão relatadas nas próximas seções.

O número total número de leituras (reads) sequenciados e alinhados nas diferentes amostras é mostrado na Tabela 4.3. Observamos que os reads alinhados no genoma de *H. salinarum* NRC-1 das réplicas do RNA total, depletados do rRNA, das linhagens foi balanceado (Tabela 4.3). O mesmo resultado foi observado para amostras do RNA total fragmentado e depletado do rRNA preparados com protocolos diferentes: TruSeq Small e TruSeq mRNA Stranded. O balanceamento de reads é indicativo que a cobertura do sequenciamento é semelhante entre as linhagens. No entanto, nas bibliotecas de Illumina RNA-Seq é comum ocorrer o desbalanceamento ou baixa profundidade de sequenciamento das amostras (Zhang et al., 2014). Este desbalanceamento ocorreu na biblioteca do tratamento TEX+ da linhagem nocaute preparada com o TruSeq Small, que foi a biblioteca que apresentou menos reads sequenciados, indicando que as análises derivadas desses dados devem ser explorados com cautela.

Tabela 4.3: Números de *reads* sequenciados e alinhados no genoma de *Halobacterium salinarum* NRC-1 nas distintas amostras das linhagens Δ ura3 (controle) e Δ VNG_RS05855 (nocaute).

Preparo de bibliotecas	Amostras	Linhagens	Réplicas ou Tratamentos	Reads Sequenciados	Reads Alinhados (%)
TruSeq Small	RNA total depletado do rRNA (sem fragmentar)	Δ ura3 (controle)	Réplica 1	1.514.184	276.746 (18%)
		Δ ura3 (controle)	Réplica 2	1.768.119	568.692 (32%)
		Δ VNG_RS05855 (nocaute)	Réplica 1	2.183.170	549.723 (25%)
		Δ VNG_RS05855 (nocaute)	Réplica 2	1.451.153	359.229 (25%)
	RNA total enriquecido com RNA primário(dRNA-seq) (sem fragmentar)	Δ ura3 (controle)	TEX+	1.930.174	533.811 (28%)
		Δ ura3 (controle)	TEX-	2.144.638	551.256 (26%)
		Δ VNG_RS05855 (nocaute)	TEX+	750.623	192.846 (26%)
		Δ VNG_RS05855 (nocaute)	TEX-	1.908.728	462.409 (24%)
	RNA total fragmentado depois da depleção do rRNA	Δ ura3 (controle)	Réplica 1	6.004.986	1.666.875 (28%)
		Δ VNG_RS05855 (nocaute)	Réplica 1	3.645.885	1.265.865 (35%)
	TruSeq mRNA Stranded	Δ ura3 (controle)	Réplica 1	2.144.638	994.853 (46%)
		Δ VNG_RS05855 (nocaute)	Réplica 1	2.241.146	1.026.762 (46%)

4.4.1 Análise dos dados de RNAseq sem fragmentação

4.4.1.1 No transcrito da linhagem nocaute a expressão de 117 genes é alterada

Para identificar os genes com expressão diferencial utilizamos as amostras que possuem mais de uma réplica biológica: RNA total depletado do rRNA (sem fragmentar). O número de réplicas é fundamental nos ensaios comparativos do sequenciamento em larga es-

cala porque permite evidenciar que as mudanças na expressão são sistemáticas, ainda quando por considerações econômicas o uso de réplicas nestes experimentos não são normalmente altas (Anders et al., 2010; Zhang et al., 2014). Um método com uma abordagem estatística adequada para identificar genes com expressão diferencial a partir de um pequeno número de réplicas é o DESeq2, amplamente utilizado e útil quando o número de reads das bibliotecas é balanceado (Michael I Love e Anders, 2014; Zhang et al., 2014; Fedorova et al., 2019; Wei et al., 2019).

Consideramos genes diferencialmente expressos na relação nocaute versus controle aqueles que apresentaram $|\log_2\text{FoldChange}| > 1$ e p-value ajustado < 0.05 . Nestas condições os genes com expressão diferencial terão mudado sua expressão pelo menos o dobro e 5% serão falsos positivos, considerações similares foram usados para estudar o transcrito de outros organismos (Barra et al., 2017; Fedorova et al., 2019; Pottanat et al., 2019). Com estes parâmetros, conseguimos identificar 117 genes diferencialmente expressos na linhagem nocaute, o que representa aproximadamente 4.5% dos genes anotados em *H. salinarum* NRC-1. Destes, 80 se apresentaram mais expressos no nocaute e 37 menos expressos. Dos 117 genes, 44% não tem uma função definida, pois estão anotados como proteínas hipotéticas e ORF-open read frame (Tabela 4.44.5).

Previamente, mediante os ensaios fenotípicos sugerimos o possível vínculo do produto proteico do gene VNG_RS05855 com o transporte, as respostas adaptativas a arsênico e os processos de obtenção de energia. Dentre os genes com expressão diferencial, conseguimos identificar que alguns deles estão anotados com estas funções. Distinguimos um total de oito genes anotados como relacionados ao transporte, deles cinco anotados com membros da família dos transportadores ABC. Genes de resposta a arsênico também estão com a expressão alterada na linhagem nocaute, assim como genes anotados como prováveis citocromos e oito possíveis reguladores transcricionais. Nove genes relacionados à produção e regulação das vesículas de gás têm expressão diferencial na linhagem nocaute.

Em conjunto, estes resultados sugerem que as sutis diferenças fenotípicas poderiam ser fruto das mudanças na expressão de genes pela interação com o produto proteico do gene VNG_RS05855; provavelmente porque poderiam ser alvos moleculares da sua suposta ação ribonucleolítica ou mediante a ação indireta sob eles.

Tabela 4.4: Genes com expressão diferencial na linhagem nocaute VNG_RS05855.

Genes	log2FC	p value adj	Anotação	Genes	log2FC	p value adj	Anotação
VNG_RS11395	6,1	9,1E-03	gas-vesicle-protein	VNG_RS11390	2,0	7,2E-06	protein-gvpD
VNG_RS04585	3,1	7,0E-04	ATPase	VNG_RS02925	2,0	5,4E-04	AbrB-family-transcriptional-regulator
VNG_RS13595	2,9	3,0E-25	hypothetical protein	VNG_RS01160	2,0	1,9E-04	MBL-fold-hydrolase
VNG_RS02935	2,7	5,4E-04	serine-protein-kinase	VNG_RS05905	1,9	2,1E-11	DUF2892-domain-containing-protein
VNG_RS02930	2,6	2,7E-04	hypothetical-protein	VNG_RS03175	1,9	9,9E-03	methyl-accepting-chemotaxis-protein
VNG_RS00725	2,5	2,7E-05	N-acetyltransferase	VNG_RS13065	1,8	2,6E-03	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS11910	2,5	3,5E-19	ORF-H0196	VNG_RS07010	1,8	3,9E-05	hypothetical-protein
VNG_RS11270	2,5	3,7E-03	hypothetical-protein	VNG_RS11690	1,8	3,1E-06	IS4-family-transposase-ISH30
VNG_RS12265	2,4	3,5E-19	ORF-H0196	VNG_RS07205	1,8	1,1E-05	Chain-amino-acid-ABC transporter
VNG_RS00205	2,4	5,4E-19	ORF-H0196	VNG_RS10625	1,8	3,6E-13	gas-vesicle-protein
VNG_RS10570	2,4	1,5E-18	ORF-H0196	VNG_RS12320	1,8	9,2E-13	gas-vesicle-protein
VNG_RS04770	2,4	1,3E-07	CBS-domain-containing-protein	VNG_RS08630	1,7	4,8E-04	MATE-family-efflux-transporter
VNG_RS10060	2,4	1,0E-12	halolysin	VNG_RS07480	1,7	1,0E-09	transcription-regulator-like-protein
VNG_RS01165	2,4	4,6E-16	MarR-family-transcriptional-regulator	VNG_RS13600	1,7	5,7E-03	hypothetical-protein
VNG_RS10660	2,3	7,8E-13	ORF-H0386	VNG_RS02055	1,7	2,7E-05	Fe-S-cluster-assembly-protein-SufD
VNG_RS09480	2,3	1,0E-07	O-acetylhomoserine	VNG_RS09905	1,7	1,6E-03	ABC-transporter-permease
VNG_RS11120	2,3	2,2E-02	hypothetical-protein	VNG_RS10615	1,7	2,4E-02	transcriptional-regulator
VNG_RS11695	2,2	2,1E-08	hypothetical-protein	VNG_RS02030	1,6	2,3E-07	hypothetical-protein
VNG_RS12355	2,2	3,7E-12	ORF-H0386	VNG_RS11730	1,6	3,0E-11	transposase
VNG_RS06995	2,2	3,1E-02	Hsp20/alpha-crystallin-family-protein	VNG_RS07200	1,6	3,8E-08	type-I-deoxyribonuclease-HsdR
VNG_RS02940	2,2	8,6E-04	serine-protein-kinase	VNG_RS10390	1,6	3,0E-02	homoserine-dehydrogenase
VNG_RS07475	2,1	6,8E-10	TrkA-family-potassium-uptake-protein	VNG_RS10585	1,6	4,1E-03	protein-gvpL
VNG_RS11725	2,1	1,0E-09	hypothetical-protein	VNG_RS12280	1,6	2,5E-03	protein-gvpL
VNG_RS07850	2,0	1,1E-02	hypothetical-protein	VNG_RS09935	1,6	1,0E-02	hypothetical-protein
VNG_RS10600	2,0	4,3E-03	protein-gvpH-1	VNG_RS10735	1,6	2,4E-02	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS10275	2,0	1,0E-07	RNA-binding-protein	VNG_RS13275	1,6	5,4E-04	hypothetical-protein
VNG_RS13360	2,0	2,9E-02	hypothetical-protein	VNG_RS12430	1,6	3,6E-02	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS03855	2,0	5,8E-07	hypothetical-protein	VNG_RS01340	1,5	2,1E-07	hypothetical-protein
VNG_RS07650	2,0	8,6E-05	hypothetical-protein	VNG_RS12325	1,5	2,5E-04	gas-vesicle-protein-GvpC

Tabela 4.5: (continuação): Genes com expressão diferencial na linhagem nocaute VNG_RS05855

Genes	log2FC	p value adj	Anotação	Genes	log2FC	p value adj	Anotação
VNG_RS01270	1,5	1,5E-05	MarR-family-transcriptional-regulator	VNG_RS12740	-8,3	3,1E-06	ORF-H1224
VNG_RS09910	1,5	1,8E-04	ABC-transporter-permease	VNG_RS12725	-8,2	6,3E-06	ORF-H1161
VNG_RS07000	1,5	1,3E-02	hypothetical-protein	VNG_RS11030	-8,1	6,5E-06	ORF-H1161
VNG_RS10630	1,4	1,2E-03	gas-vesicle-protein-GvpC	VNG_RS12730	-8,1	6,5E-06	ORF-H1186
VNG_RS07015	1,4	5,2E-04	hypothetical-protein	VNG_RS11035	-8,1	6,5E-06	ORF-H1186
VNG_RS07230	1,4	2,2E-03	hypothetical-protein	VNG_RS12800	-8,0	1,3E-05	ORF-H1364
VNG_RS11555	1,3	1,1E-04	catalase-peroxidase	VNG_RS11755	-7,8	2,1E-05	hypothetical-protein
VNG_RS02025	1,3	1,3E-02	hypothetical-protein	VNG_RS12900	-7,1	4,4E-04	ORF-H1517
VNG_RS10610	1,3	9,8E-03	protein-gvpF	VNG_RS12815	-7,0	1,7E-02	pyridoxamine-5'-phosphate
VNG_RS10640	1,3	2,5E-02	protein-gvpO-1	VNG_RS12790	-6,9	8,3E-04	ORF-H1336
VNG_RS12305	1,3	1,5E-02	protein-gvpF	VNG_RS12760	-6,9	8,8E-04	hypothetical-protein
VNG_RS08810	1,3	3,5E-02	phosphoenolpyruvate-carboxylase	VNG_RS13840	-6,6	2,4E-03	ORF-H1537
VNG_RS13805	1,2	1,1E-03	hypothetical-protein	VNG_RS12910	-6,4	3,7E-03	DNA-binding-protein
VNG_RS12335	1,2	3,5E-02	protein-gvpO-1	VNG_RS13820	-6,4	3,9E-03	hypothetical-protein
VNG_RS01815	1,2	2,0E-02	aldo/keto-reductase	VNG_RS13830	-6,1	1,1E-02	transcriptional-regulator
VNG_RS10620	1,2	3,3E-03	protein-gvpD	VNG_RS12830	-6,0	1,5E-02	ORF-H1420
VNG_RS13470	1,1	4,8E-03	hypothetical-protein	VNG_RS12855	-5,9	2,2E-02	arsenic-transporting-ATPase
VNG_RS02950	1,1	1,3E-02	PadR-family-transcriptional-regulator	VNG_RS13625	-5,8	2,5E-02	hypothetical-protein
VNG_RS05385	1,1	1,9E-02	hypothetical-protein	VNG_RS12735	-5,5	5,0E-02	transcription-factor
VNG_RS04125	1,0	2,4E-03	glycosyltransferase-WcaL	VNG_RS12880	-4,9	1,3E-03	hypothetical-protein
VNG_RS07785	1,0	6,1E-03	DNA-binding-protein	VNG_RS04715	-2,9	8,4E-05	hypothetical-protein
VNG_RS09555	1,0	2,8E-03	DNA-starvation/stationary-phase	VNG_RS05265	-2,5	4,7E-03	EamA-family-transporter
VNG_RS13630	-10,6	7,0E-11	hypothetical-protein	VNG_RS10465	-1,6	1,0E-02	archaeosortase-A
VNG_RS12750	-9,5	1,0E-09	ORF-H1238	VNG_RS04530	-1,5	3,7E-02	nucleoside-diphosphate-kinase
VNG_RS12865	-9,5	1,5E-08	ArsC-family-transcriptional-regulator	VNG_RS01775	-1,4	1,4E-05	Phosphate-ABC-transporter
VNG_RS12920	-9,3	3,5E-08	ORF-H1549	VNG_RS11450	-1,4	3,5E-02	cell-division-protein
VNG_RS12765	-8,8	3,4E-07	ORF-H1292	VNG_RS03605	-1,2	5,6E-05	hypothetical-protein
VNG_RS12745	-8,7	5,4E-07	ORF-H1232	VNG_RS11195	-1,1	8,2E-05	universal-stress-protein
VNG_RS12755	-8,4	2,3E-07	ORF-H1243	VNG_RS09710	-1,1	1,5E-04	Phosphate-ABC-transporter-PstC
				VNG_RS06830	-1,1	4,1E-02	sensory-rhodopsin-2

4.4.1.2 Processos biológicos relacionados aos genes com expressão diferencial

Uma variedade de resultados tem mostrado que as RNases desempenham papéis críticos na regulação de processos biológicos específicos, na sua maioria dependentes do ambiente celular: mudanças na concentração de ferro, controle do ciclo celular, transições entre estados fisiológicos (Lee, Bernstein e Cohen, 2002; Evguenieva-Hackenberg e Klug, 2011; Wurtmann et al., 2014). Para identificar quais processos biológicos estão afetados pela ausência do gene VNG_RS05855, nós focamos em reconhecer aqueles que estão enriquecidos nos genes com expressão diferencial.

O estudo do enriquecimento de processos biológicos foi realizado utilizando os termos GO (Gene Ontology). Para validar nossos resultados utilizamos o teste exato de Fisher (Lewin e Grieve, 2006; Huang, Sherman e Lempicki, 2009; Glass e Girvan, 2014). Nós focamos em identificar os 20 processos biológicos mais representados.

Observamos que o processo biológico mais enriquecido é a organização das vesículas de gás. Em níveis similares de significância aparecem os processos biológicos relacionados a resposta a arsênico e a outros estímulos. Também enriquecidos estão os processos biológicos relacionados com mecanismos de reparo ao DNA por recombinação (Tabela 4.6).

Tabela 4.6: Termos *GO* enriquecidos nos genes com expressão diferencial (DESeq2) na linhagem nocaute VNG_RS05855

Termo GO	Descrição	Anotados	Significantes	Esperado	p-valor
GO:0031412	gas vesicle organization	12	4	0,42	5,2E-04
GO:0046685	response to arsenic-containing substance	6	3	0,21	7,4E-04
GO:0050896	response to stimulus	166	14	5,8	8,0E-04
GO:0006310	DNA recombination	43	6	1,5	2,9E-03
GO:0016043	cellular component organization	91	9	3,18	3,0E-03
GO:0009314	response to radiation	10	3	0,35	4,0E-03
GO:0065004	protein-DNA complex assembly	10	3	0,35	4,0E-03
GO:0071824	protein-DNA complex subunit organization	10	3	0,35	4,0E-03
GO:0009628	response to abiotic stimulus	11	3	0,38	5,4E-03
GO:0000724	DSB repair via homologous recombination	4	2	0,14	6,8E-03
GO:0000725	recombinational repair	4	2	0,14	6,8E-03
GO:0000730	DNA recombinase assembly	4	2	0,14	6,8E-03
GO:0010212	response to ionizing radiation	4	2	0,14	6,8E-03
GO:0042148	strand invasion	4	2	0,14	6,8E-03
GO:0045003	DSB repair via synthesis	4	2	0,14	6,8E-03
GO:0090735	DNA repair complex assembly	4	2	0,14	6,8E-03
GO:0071840	cellular component organization(biogenesis)	105	9	3,67	7,9E-03
GO:0042221	response to chemical	72	7	2,52	1,0E-02
GO:0006312	mitotic recombination	5	2	0,17	1,1E-02
GO:0009605	response to external stimulus	46	5	1,61	1,9E-02

Em *H. salinarum* NRC-1 há evidências de que a formação de vesículas de gás diminui durante o dia para evitar a exposição a radiações que afetem o DNA (Pfeifer, 2012). Perante o incremento na expressão de genes das vesículas de gás, é possível que apareçam sinais que ativem mecanismos de proteção às radiações, como os mecanismos de recombinação. O interessante é que a formação de vesículas de gás é reprimida com inibidores de girases (DasSarma, S, 1992), enzimas chaves nos processos de recombinação, demonstrando que há sinais comuns entre ambos eventos. Nesse sentido, nós propomos que um possível fator nessa interconexão seja VNG_RS05855, mediando mecanismos pós-transcricionais para a regulação de ambos eventos. Estudos recentes em eucariotos e bactérias já sugeriram que processos pós-transcricionais contribuem nas complexas respostas de reparo ao DNA, em alguns casos vinculados a pequenos RNA (sRNA) e sem a mediação de proteínas chaperonas (McKay, 2013; Fröhlich, Förstner e Gitai, 2018). Outra hipótese é a característica das regiões genômicas onde se insere os cluster das vesículas de gás. Elas estão na vizinhança de sequências de inserção (IS) (Jones, Young e DasSarma, 1991), e a transcrição ativa destas regiões poderiam promover sinais moleculares para ativar a recombinação por transposição justificando o fato deste processo molecular enriquecido na linhagem nocaute. Contudo, eventos de recombinação e eventuais regulações pós-transcricionais são poucos conhecidos em *H. salinarum* NRC-1.

A seguir, discutiremos os grupos de genes com expressões diferenciais associadas aos três principais processos biológicos que também foram explorados mediante ensaios fenotípicos e que poderiam auxiliar a caracterizar o gene VNG_RS05855.

4.4.1.2.1 Organização de vesículas de gás

Em *H. salinarum* NRC-1, um total de 14 genes, conhecidos como *gvp*, estão envolvidos na formação de vesículas de gás; eles se dispõem em dois grupos orientados de forma oposta, *gvpACNO* e *gvpDEFGHIJKLM*, que estão espalhados em várias cópias nos elementos extracromossômicos de *H. salinarum* NRC-1. Nossos resultados mostram a expressão aumentada na linhagem nocaute dos genes *gvpA* (codificados por VNG_RS10625, VNG_RS12320, VNG_RS11395) e *gvpD* (VNG_RS10620, VNG_RS11390). O gene *gvpA* codifica para o maior componente estrutural das vesículas de gás e o gene *gvpD* para o repressor negativo da transcrição dos membros destes operons (Pfeifer, 2015). Além deles, outros genes estruturais

também estão com expressão aumentada na linhagem nocaute e podem ser consultados na Tabela 4.4.

O incremento na transcrição dos genes *gvpA* sugere a acumulação de vesículas de gás na linhagem nocaute em relação à linhagem controle. No entanto, o aumento do nível da expressão do repressor *gvpD* poderia indicar o contrário. As microscopias das células das linhagens controle e nocaute não mostraram grandes diferenças na quantidade de vesículas de gás (Figura 4.18).

Se os níveis de mRNA destes genes estão alterados e não se conseguem observar diferenças na quantidade de proteínas derivadas delas, uma possibilidade é que os eventos de tradução sejam afetados pela ausência do gene VNG_RS05855. Propomos que possivelmente na linhagem nocaute ocorra o incremento na estabilidade de elementos regulatórios que impeçam a tradução dos transcritos do operon das vesículas de gás.

4.4.1.2.2 Respostas a arsênico

Entre os genes com expressão diferencial identificamos que os que respondem ao metalóide arsênico tem a expressão diminuída na linhagem nocaute em relação ao controle. Eles são os genes VNG_RS12865, VNG_RS12855 e VNG_RS13830, anotados como redutase de arsênico (*arsC*), ATPase transportadora de arsênico (*arsA*) e ao regulador da resposta ao metalóide (*arsR*), respectivamente. A expressão diminuída destes genes pode ser a confirmação do fenótipo de sensibilidade da linhagem nocaute ao arsênico (Figura 4.7).

Em investigações da funcionalidade do conjunto de genes em *H. salinarum* NRC-1 de resposta a arsênico se concluiu que na sua ausência as células são susceptíveis. No entanto, se comprovou que os supostos repressores, *arsR* e *arsD*, não influenciam os níveis de resistência a este metalóide. Também mostraram que a expressão do *arsC* na presença do arsênico é baixa, e somente a expressão do gene *arsA* mostrou influenciar a resistência a este metalóide neste organismo (Gejiao Wang, Sean P. Kennedy, Sabeena Fasiludeen, Christopher Rensing e DasSarma, 2006). Em nossos resultados a diminuição da expressão do gene *arsA* na linhagem nocaute poderia indicar a sua susceptibilidade perante arsênico.

Em *E. coli*, Joshua e colaboradores demonstraram experimentalmente que o decaimento diferencial dos mRNA contidos na unidade policistônica do operon de resposta a arsê-

nico é essencial para garantir a resposta eficiente na presença deste metalóide (Owolabi e Rosen, 1990). Nós sugerimos, que em *H. salinarum* NRC-1 o processamento pós-transcricional também seja chave para garantir a resposta adaptativa ao metalóide e com a eventual participação do VNG_RS05855.

4.4.1.2.3 Respostas a estímulos

Uma variedade de genes poderiam estar relacionados com os processos biológicos de respostas a estímulos. Ajustes fisiológicos imediatos para manter a homeostase poderiam estar vinculados a reguladores transcricionais e até mesmo mecanismos de transporte (Baliga et al., 2002; Kaur et al., 2006; Wurtmann et al., 2014). Pela ampla diversidade, decidimos concentrar nossos esforços em estudar, brevemente, os processos mais representados nos genes com expressão diferencial.

4.4.1.2.3.1 Transportadores

A família de transportadores ABC (ATP-binding cassette) estão presentes nos três domínios da vida, e estão envolvidos em processos que mediam desde a resistência a metais até a osmo-regulação celular (Tomii e Kanehisa, 1998; Srivastava e Kowshik, 2013). Pelo menos 27 membros da superfamília de transportadores estão presentes em *H. salinarum* NRC-1 (Ng et al., 2000). Destes cinco tem uma expressão diferencial na linhagem nocaute em relação a linhagem controle: três com expressão aumentada e dois com a expressão diminuída.

Os genes com expressão aumentada na linhagem nocaute são VNG_RS09905 (dppB2), VNG_RS09910 (dppC2) e VNG_RS07205. Os dois primeiros genes codificam permeases, capazes de internalizar dipeptídeos na célula, e o terceiro uma proteína de ligação a aminoácidos. Os transportadores Dpp estão envolvidos, além do transporte de peptídeos, com o transporte de níquel e a formação do biofilmes em *H. salinarum* NRC-1, e nas bactérias se associam com a expressão de mecanismos de defesa e a adaptação às mudanças ambientais (Völkel, Fröls e Pfeifer, 2018). As mudanças na expressão destes genes poderiam ter algum efeito na manutenção das propriedades osmóticas das células, e ter alguma relação com a funcionalidade descrita em bactérias, onde a linhagem apresentaria maior facilidade para a adaptação às mudanças ambientais. Contudo, observamos, mediante as diferenças de OD_{600nm}, que a linhagem nocaute

não é mais resistente que a linhagem controle diante das mudanças nas concentrações de sais do meio e que frente às perturbações originadas por metais de transição, a resistência da linhagem nocaute aparece só nas menores concentrações avaliadas (Figura 4.10).

Os genes com expressão diminuída VNG_RS01775 (pstS2) e VNG_RS09710 (pstC1) também pertencem a família de transportadores ABC. Eles formam parte de operons distintos (pst2 e pst1) com a função no transporte de fosfato e codificam, respectivamente, para uma proteína de captação de fosfato extracelular e para uma permease. Frente a alguma perturbação ambiental, a diminuição da expressão destes genes poderia comprometer a internalização do fosfato inorgânico e com isso afetar os processos de resistência a metais pesados e de adaptação a mudanças osmóticas (Orell et al., 2012). Esta situação não consegue justificar a sutil resistência da linhagem nocaute frente a maioria das concentrações dos metais de transição, porém poderia explicar a sensibilidade da linhagem nocaute diante das mudanças de salinidade, visto os valores negativos das diferenças OD_{600nm} (Figura 4.8).

Além dos transportadores ABC, identificamos outro gene relacionado ao transporte, VNG_RS07475 que apresenta na sua sequência domínios TrKA, associados com a captação de potássio do meio extracelular (Ng et al., 2000). Em *E. coli* a movimentação de K⁺ por estes transportadores é influenciada pela relação NAD⁺:NADH, sendo que a primeira espécie é capaz de unir-se ao transportador e facilitar a sua movimentação (Schlosser et al., 1993), fato que ajuda a explicar o por que estes transportadores funcionam de maneira diferente nas condições aeróbicas e anaeróbicas (Bagramyan e Martirosov, 1989). Em *H. salinarum* NRC-1 foi sugerida a degradação ribonucleolítica deste transportador nas condições aeróbicas (Wurtmann et al., 2014). O aumento na expressão deste transportador na linhagem nocaute do gene VNG_RS05855 faz sugerir que o seu produto proteico esteja vinculado ao possível processamento pós-transcricional deste transportador. A ação em conjunto desta possível RNase com a RNase-VNG2099 já foi sugerida previamente nas análises de correlação (Figura 4.3 III), concluindo a ação redundante e inespecífica das RNases.

Outros transportadores com expressão diferencial são os que participam na saída da célula de compostos do entorno celular como os genes VNG_RS08630, VNG_RS05265 e VNG_RS05265 relacionados à família dos transportadores MATE (Multidrug And Toxic Com-

pound Extrusion) e a família de transportadores EamA, respectivamente (Franke et al., 2003; Bay, Rommens e Turner, 2008; Kuroda e Tsuchiya, 2009; Västermark et al., 2011). Porém, nas archaeas as informações que se tem do conjunto destes transportadores são limitadas (Kira S. Makarova, Michael Y. Galperin e National, 2016).

4.4.1.2.3.2 Reguladores transcricionais

Dois genes (VNG_RS01165 e VNG_RS01270) anotados como membros da família de reguladores MarR aparecem com expressão diferencial na linhagem nocaute. A família destas proteínas são reguladores da expressão de genes que codificam bombas de efluxos como os transportadores MATE(Grove, 2017). Porém, em bactérias e arqueias estas proteínas estão associadas a regulação de outros genes que também são críticos nas respostas ambientais. A maioria das proteínas MarR são repressores transcricionais, porém recentemente foi descoberto que podem funcionar como ativadores e ser parte de operons que estão sob a sua própria regulação, promovendo a transcrição diferencial pela união a promotores divergentes localizados nas regiões intergênicas (Wilkinson e Grove, 2006).

Outros genes anotados com função reguladora com expressão diferente entre as linhagens são os genes VNG_RS02950 e VNG_RS02925. O primeiro pertence a família de reguladores transcricionais PadR, relacionadas com o controle negativo da expressão da descarboxilase de ácido fenólico que funcionam como sensores ambientais (Park et al., 2017). O segundo, pertence a família de proteínas reguladoras AbrB. Em bactérias, membros desta família atuam tanto como ativadores ou repressores da expressão de genes relacionados com a transição entre o crescimento vegetativo e o início da fase estacionária e da esporulação (Strauch et al., 2006; Daniel M. Sullivan, Benjamin G. Bobay, Douglas J. Kojetin, Richele J. Thompson, Mark Rance, Mark A. Strauch e Cavanagh, 2009). Os conhecimentos dos aspectos bioquímicos e biofísicos que caracterizam o mecanismo regulador da sua atividade são escassos(Lee, Kim e Hong, 2017).

A limitada compreensão do papel das proteínas reguladoras na expressão de genes relacionados aos transportadores em *H. salinarum* NRC-1 deixa aberta a questão dos transcritos com expressões diferenciais serem alvos diretos do produto proteico do gene VNG_RS05855, ou consequência das mudanças originadas pelos reguladores transcricionais que poderiam, ou

não, estar sob seu controle.

4.4.1.2.3.3 Outros genes

Nesta categoria de genes que poderiam estar relacionados a respostas a estímulos decidimos incluir outros quatro genes cuja função foi explorada mediante ensaios fenotípicos mas não estão associados ao transporte ou elementos regulatórios da transcrição, apesar da sua expressão estar relacionada com habilidades adaptativas nas células.

Os genes VNG_RS10735, VNG_RS12430 e VNG_RS13065 mais expressos na linhagem nocaute em relação à linhagem controle, são cópias localizadas nos elementos extracromossômicos de *H. salinarum* NRC-1 que codificam para a citocromo-d-ubiquinol-oxidase subunit II (cydB2). Localizados upstream a cada um deles estão genes que codificam para citocromo-d-ubiquinol-oxidase subunit I (cydA2), os pares organizados em forma de operons constituem o complexo IV da via da fosforilação oxidativa, o citocromo bd (Klein et al., 2005).

A principal função do citocromo bd, estritamente encontrada em procariotos, é a redução do oxigênio molecular a água mediante a oxidação do ubiquinol a ubiquinona. Nesta reação se gera uma força próton motriz por transferência de carga vetorial sem bombeamento real de prótons e ocorre fundamentalmente quando a célula cresce com suprimento limitado de oxigênio (Puustinen et al., 1991). Este último fato se junta a outras funcionalidades descritas como: apoiar o crescimento fotossintético anaeróbico, mecanismos de desintoxicação do peróxido de hidrogênio e degradação de compostos aromáticos (Vitaliy B. Borisova, Robert B. Gennissb, James Hempb, Verkhovskyc e Aelozersky, 2011). Contudo, o aumento na sua expressão não representou uma vantagem para a linhagem nocaute frente às perturbações provocadas por agentes oxidantes. Nossa hipótese era que o fato de ser cyB2 membro dos citocromos influenciaria as diferenças no pH das linhagens (Figura 4.14 B) por garantir transferência de carga vetorial. No entanto, em *E. coli* este gene mostrou ter uma maior relação com processos que conferem competências adaptativas as células do que com mudanças de pH (Puustinen et al., 1991).

Outro gene associado a respostas a estímulos, porém com expressão diminuída na linhagem nocaute, é VNG_RS06830. Este gene codifica para o sensor II da rodopsina (SRII) e tem expressão diminuída na linhagem nocaute em relação à linhagem controle. O SRII fun-

ciona exclusivamente como um receptor repelente da luz azul que permite que o organismo procure o escuro quando a respiração e os nutrientes são abundantes. Esta proteína é produzida constitutivamente pela célula, no entanto, ela tem uma baixa concentração (Schäfer, Engelhard e Müller, 1999; Person et al., 2005; Inoue, Tsukamoto e Sudo, 2014). Na linhagem nocaute nós não conseguimos distinguir diferenças entre os picos de emissão desta proteína com a linhagem controle nos ensaios realizados por espectroscopia Raman (Figura 4.17 F). Na literatura não encontramos nenhuma referência ao processamento pós-transcricional para atingir a funcionalidade deste fotossensor. Contudo, os valores da sua expressão poderiam sugerir algum tipo de relação com o produto proteico do gene VNG_RS05855, ausente na linhagem nocaute.

4.4.1.3 Estruturas dos sinais de expressão gênica

Além de fornecer resultados a respeito dos níveis de expressão gênica, os dados de RNA-seq trazem outra informação relevante para o entendimento do transcrito: a estrutura do sinal de expressão gênica. Uma vez que os reads são gerados e mapeados ao genoma referência, é possível contabilizar a cobertura em cada posição do genoma e gerar um sinal de expressão para todas as posições mapeadas. Com essa abordagem, o RNA-seq vem auxiliando na detecção de regiões alvos da atividade de RNases pela comparação da cobertura de reads entre linhagens controle e nocaute de enzimas com esta atividade (Yanjie Chao, 2017).

4.4.1.3.1 Perfil dos sinais de expressão dos genes associados a processos biológicos

Com o objetivo de identificar possíveis sinais associados ao processamento pós-transcricional, analisamos o perfil dos sinais de expressão para os genes relacionadas aos principais processos biológicos afetados pela ausência na linhagem nocaute do VNG_RS05855.

Para o gene *gvpA* do operon de vesículas de gás, a cobertura de reads na fita oposta, especificamente na sua extremidade 5' mostra que a intensidade do sinal de expressão é superior na linhagem nocaute em relação ao controle. Para esta região foi proposta a existência de um RNA antisense que poderia bloquear a tradução do gene *gvpA* (Almeida et al., 2019). Um mecanismo similar de regulação foi apresentado por Krüger e col tanto nas extremidades 5' e 3' do gene *gvpD* (Krüger e Pfeifer, 1996). Porém, nessas regiões do gene *gvpD* não conseguimos

distinguir sinais na cobertura de reads que sejam diferentes entre as linhagens. Consideramos que o possível modo da atuação da proteína codificada no gene VNG_RS05855 se relacione com o controle da estabilidade de RNA antisense que possam regular a expressão de genes do operon gvp (Figura 4.22).

Figura 4.22: Cobertura de *reads* das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* em “*” descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. Gene sinalizado em negrito está com expressão aumentada na linhagem nocaute segundo as análises do DESeq2 e representam membros do operon de vesículas de gás. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019) e as caixas vermelhas os resultados da interação da proteína Lsm com RNA em ensaios de RIP-seq (Zaramela, 2015).

Observamos os sinais de cobertura de reads para os membros do operon de resposta a arsênio (VNG_RS12855, VNG_RS13830 e VNG_RS12865), além de apresentar maiores sinais de expressão na linhagem controle, as diferenças na cobertura de reads entre os genes que o constituem pode indicar diferenças na estabilidade dos transcritos da unidade policistônica, similar ao sugerido para o operon de arsênio de *E. coli* (Owolabi e Rosen, 1990). Na ausência do gene VNG_RS05855, a expressão dos membros deste operon é afetada, sugerindo que ele esteja relacionado a fatores que incrementam a estabilidade ou expressão deles (Figura 4.23).

Figura 4.23: Cobertura de *reads* das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* em “*” descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. Gene sinalizado em negrita está com expressão diminuída na linhagem nocaute segundo as análises do DESeq2 e representam membros do operon de resposta a arsênio. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019) e as caixas vermelhas os resultados da interação da proteína Lsm com RNA em ensaios de RIP-seq (Zaramela, 2015).

Observamos também o perfil de cobertura de reads dos transportadores com expressão aumentada na linhagem nocaute, dppB2 e dppC2, e os que apresentam expressão diminuída, pstS2 e pstC1. Os genes dppB2 e dppC2 formam parte de o mesmo operon (Price et al., 2005) e estão localizados na extremidade 3' do transcrito policistrônico, o aumento na sua expressão poderia indicar o aumento na estabilidade desta extremidade diante a ausência do VNG_RS05855. Reanálises da cobertura de reads das linhagens nocaute das exoribonucleases RNases II, PNPa-ses e RNase R de *E. coli* (Pobre e Arraiano, 2015) diante de seus transcritos alvos apresentam perfis de cobertura de reads similares aos observados nos transcritos dos transportadores DppB2 e DppC2 na linhagem nocaute. Para os genes pstS2 e pstC1, também relacionados ao transporte, a maior expressão é na linhagem controle que na nocaute, porém a localização destes genes no operon é mais próxima ao início da transcrição da unidade policistrônica. Neste sentido, sugerimos que as diferenças na expressão destes genes na ausência do VNG_RS05855, possivelmente,

esteja vinculada com a sua localização nas unidades poliestrônicas (Figura 4.24).

Figura 4.24: Cobertura de *reads* das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* em “*” descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. A) Genes sinalizados em negrito estão com expressão aumentada na linhagem nocaute segundo as análises do DESeq2 e representam membros de transportadores ABC capazes de transportar dipéptidos. B) Gene sinalizado em negrito está com expressão diminuída na linhagem nocaute segundo as análises do DESeq2 e representam membro dos transportadores ABC capazes de transportar fosfato inorgânico. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019).

Por último, focamos no perfil de expressão do gene VNG_RS01165, que pertence a família de reguladores transcricionais MarR. Estes reguladores geralmente formam parte de operons que estão sob a sua própria regulação e promovem a transcrição diferencial de seus membros pela união a promotores divergentes localizados nas regiões intergênicas (Wilkinson e Grove, 2006). Desta maneira, e pela presença de promotores alternativos em *H. salinarum* NRC-1 (Koide et al., 2009), conseguimos justificar os incrementos no perfil de expressão do gene adjacente a ele, VNG_RS01160 (Figura 4.25) Embora, seus aumentos possam ser relacionados com a regulação transcricional, a ausência de uma possível RNase também poderia justificar os incrementos observados pelo aumento da estabilidade do transcrito poliestrônico.

Figura 4.25: Cobertura de *reads* das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* em “*” descrevem a média das contagens das réplicas por linhagens. Caixas laranjas representam genes na fita *reverse*. Genes sinalizados em negrita estão com expressão aumentada na linhagem nocaute segundo as análises do DESeq2.

As análises do perfil da cobertura de reads de genes com expressão diferencial na linhagem nocaute permitiu identificar sinais que sugerem a participação do produto proteico do gene VNG_RS05855 na regulação pós-transcricional. A possível RNase pode deixar rastros do processamento desigual pela distribuição heterogênea de potenciais sítios de processamento entre as linhagens controle e nocaute. Nós focamos em identificar esses sítios mediante a comparação da cobertura de reads em todas as posições dos genomas utilizando abordagens computacionais. Com esse propósito, exploramos o sinal do transcrito dos dados do sequenciamento das amostras RNA total depletado de rRNA (sem fragmentar) preparadas com o protocolo TruSeq Small, as mesmas que utilizadas na identificação dos genes com expressão diferencial entre as linhagens.

4.4.1.3.2 Rastros do processamento desigual entre as linhagens por diferenças nos sinais de expressão

Explorar os sinais de descontinuidades ou rupturas do perfil de cobertura de reads, obtidas a partir do sequenciamento de amostras sem fragmentar e preparadas com o protocolo que conserva especificidade de fita (TruSeq Small), permite associar-lhas a potenciais sítios de processamento ribonucleolítico. Para isto, consideramos que as descontinuidades nos sinais dos transcritos não são originadas pela fragmentação experimental. Elas poderiam ser o resultado da

acumulação de reads entre extremidades adjacentes do transcrito, promovidos por cortes endoribonucleolíticos sem a posterior ação de exoribonuclease. As extremidades geradas formariam variantes do transcrito. Por outro lado, as possibilidades de que essas extremidades adjacentes correspondam com inícios de transcritos alternativos e finais de outros transcritos são baixas, em função da ampla formação de isoformas durante a terminação da transcrição em Archaea (Dar et al., 2016).

Nosso primeiro passo foi identificar as posições adjacentes no genoma de cada linhagem que apresentassem o acúmulo de reads. Após identificar essas posições, buscamos as que estão presentes na linhagem controle e ausentes na linhagem nocaute (Figura 3.2). A inexistência das posições na linhagem nocaute sugere que podem ser as potenciais posições alvos do processamento da atividade ribonuclease de produto proteico do gene VNG_RS05855.

Com esta abordagem identificamos que o número de possíveis posições de processamento entre as linhagens não é marcadamente diferente, apesar da ausência de uma suposta RNase na linhagem nocaute (Figura 4.26). Este resultado pode ser relacionado a atividade redundante deste tipo de enzimas. Outra provável explicação é a instabilidade de extremidades originadas por cortes endoribonucleolíticos que são rapidamente processados por exoribonucleases.

Figura 4.26: Sítios de processamento nos transcritos dos replicons das linhagens controle (Δ ura3) e nocaute (Δ VNG_RS05855) pelo procedimento desenvolvido “in-house” a partir dos sinais da cobertura de *reads* obtidas do sequenciamento de amostras de RNA total depletado de RNA ribossomal preparadas com o protocolo TruSeq Small.

Do total dos sítios de processamento, 191 podem ser consideradas como potenciais alvos do produto proteico do gene VNG_RS05855 por estarem só presente na linhagem controle (Figura 4.27 A). Estes sítios têm uma distribuição desigual entre os diferentes replicons e análises da densidade da sua distribuição, calculada como a relação entre o número de posições de processamento e o tamanho do replicon, revelam, além do processamento diferencial nos replicons em *H. salinarum* NRC-1, que a possível proteína codificada no gene VNG_RS05855 processa o plasmídeo pNRC-100 com mais intensidade (Figura 4.27 B).

Figura 4.27: Sítios de processamento dos transcritos (TPS-transcriptional processing sites), identificados pelo procedimento desenvolvido “in-house” a partir dos sinais da cobertura de *reads* obtidas do sequenciamento de amostras de RNA total depletado de RNA ribossomal sem fragmentar, associados a atividade do produto proteico do gene VNG_RS05855. A) Número de TPS identificados em cada replicon. A representação não está realizada em escala. B) Densidade de TPS em cada replicon.

Nenhum dos sítios de processamento pós-transcricional identificados com esta abordagem estão localizados nos operons envolvidos no transporte e na produção de vesículas de gás. No entanto, foram detectadas posições de processamento no operon de arsênio, porém as posições não se localizam dentro das regiões codificadoras dos genes com expressão diferencial na linhagem nocaute.

Do total de posições de processamento, 96 estão localizadas nos rRNA 23S, 16S e 5S e 21 localizadas em 10 tRNAs, incluindo o tRNA-Trp, o único em *H. salinarum* NRC-1 que possui íntrons (Figura 4.28).

Figura 4.28: Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 representados em triângulos cinzas localizados maioritariamente nos A) rRNA 16S e 23S e B) tRNA-Trp, onde a área sombreada representa a região gênica que se corresponde com o intron dessa sequência. Todos os genes localizados na fita *forward*. Os sítios de processamento foram determinados segundo análises feitas “*in house*” com os dados da média da cobertura de *reads* por réplicas preparadas com o protocolo TruSeq-Small sem fragmentar das linhagens controle (verde) e nocaute (vermelho). Caixas vermelhas os resultados dos ensaios de RIP-seq (Zaramela, 2015).

Os possíveis sítios de processamento identificados estão em regiões onde se propõe que outras RNases atuem. Por exemplo, o splicing no tRNA-Trp de *H. salinarum* NRC-1 é proposto que seja realizado pelo produto proteico codificado no gene VNG_RS08615, anotado no RefSeq como a tRNA splicing endonuclease. Da mesma maneira a presença de sítios alvos dessa proteína nos tRNAs. Nas regiões dos rRNA, a atividade pós-transcricional é sugerida que seja feita por um grupo RNases (Clouet-d’Orval, 2017). Em conjunto, estes argumentos podem indicar a provável coordenação entre a possível RNase codificada no gene VNG_RS05855 com as RNases que possuem funções específicas nas células.

4.4.2 Análise dos dados de RNAseq com fragmentação

As bibliotecas para identificar os genes com expressão diferencial descritas na seção 3.5.4 foram preparadas com o protocolo TruSeq Small a partir de amostras sem fragmentar do RNA total, depletadas de rRNA. Este protocolo seleciona fragmentos entre 150 e 600 pb para o sequenciamento. Fragmentos com tamanhos superiores afetam o sequenciamento das

amostras. Após identificar, com o DESeq2, os genes diferencialmente expressos analisamos seus comprimentos e percebemos que o 62% deles superam os 600 pb (52 genes com expressão aumentada e 21 com expressão diminuída). Em amostras sem fragmentar e nas que foram selecionadas para o sequenciamento fragmentos menores a 600pb, a presença de transcritos com comprimentos maiores sugerem que neles ocorram processos pós-transcricionais, os que poderiam relacionar-se a possíveis eventos de degradação ou maturação.

Para estudar se este cenário muda ao examinar fragmentos maiores do que 600 pb realizamos o estudo do RNA total depletado do RNA ribossomal com o protocolo TruSeq-Stranded. Neste protocolo não se realiza uma seleção por tamanho nas amostras do RNA total depletado de rRNA, no entanto há etapas de fragmentação das amostras antes de realizar o sequenciamento para homogeneizar o tamanho do material genético a ser sequenciadas para favorecer este processo.

Também foram sequenciadas amostras de RNA total, depletadas do rRNA e fragmentadas preparadas com o protocolo TruSeq Small, nas que se realizou a seleção de fragmentos entre 150 e 600 pb. Com esta abordagem também conseguimos estudar fragmentos maiores a 600pb porque apesar de realizar a seleção por tamanho, é incorporada na preparação das amostras passos de fragmentação. O trabalho com este protocolo ainda tem a potencialidade de conservar melhor a informação da especificidade da fita e permitir enriquecer a cobertura de reads nos transcritos menores do genoma de *H. salinarum* NRC-1.

Resumindo, o intuito destes dois tratamentos foi similar: com as etapas de fragmentação, verificar as mudanças no transcrito das linhagens controle e nocaute sem o viés da amostragem de RNAs.

4.4.2.1 Papel de VNG_RS05855 no processamento de transcritos longos relacionados a diversos processos biológicos

Para estimar a possível influência do produto proteico do gene VNG_R05855, identificamos primeiro os genes com expressão diferencial entre as linhagens nocaute e controle nas bibliotecas preparadas a partir de amostras fragmentadas com os protocolos TruSeq Stranded e TruSeq Small. Porém, por só utilizar uma réplica das linhagens na construção destas bibliotecas não conseguimos realizar as análises da expressão diferencial com o DESeq2. As análises

foram realizadas utilizando os gráficos de MA-plot, que representam o $\log_2$ da média das contagens de reads normalizadas (A) versus o $\log_2$ da razão da contagem de reads (M) em cada gene das linhagens da relação nocaute vs controle. Identificamos aqueles com expressão diferencial se os valores são $|M(\log_2FC)| > 1$ e $A(\log_2 \text{ média das contagens}) > 5$ (Pobre e Arraiano, 2015). Após identificar os genes com expressão diferencial entre as linhagens, o segundo passo foi verificar seus tamanhos.

A partir das amostras de RNA total depletadas de rRNA, preparadas com o protocolo TruSeq-Stranded conseguimos identificar 401 genes com expressão diferencial entre as linhagens nocaute e controle: 245 com a expressão aumentada e 157 com a expressão diminuída (Figura 4.29 A). Dos 401 genes, 223 tem um comprimento maior que 600 nt (139 com expressão aumentada e 84 diminuída) e deles 16 são identificados com expressão diferencial entre as linhagens nocaute e controle pelo DESeq2. Destes genes em comuns, 12 tem comprimento maior do que 600 nt.

Já com o protocolo TruSeq-Small para estudar o RNA total fragmentado depletado do rRNA, conseguimos identificar 102 genes com expressão diferencial na linhagem nocaute em relação à linhagem controle (90 com expressão aumentada e 12 com expressão diminuída) (Figura 4.29 B). Deste conjunto, 52 genes têm um comprimento maior que 600 nt e 9 genes são identificados também pelo DESeq2. Dos 9 genes em comuns, 8 tem um comprimento maior que 600 nt.

Com o protocolo sem seleção por tamanho, o protocolo TruSeq Stranded mostrou o maior número de genes com expressão diferencial (Figura 4.29 A).

Ao estudar o comprimento dos conjuntos de genes com expressão diferencial, identificados pelos gráficos de MA-plot, observamos que mais da metade supera os 600 pb (Figura 4.30), independentemente das amostras serem fragmentadas, incluindo ou não a seleção por tamanho de material a ser sequenciado. O resultado nos leva a sugerir o processamento pós-transcricional é intenso em transcritos longos. Em *Sulfolobus* quanto mais longo é o transcrito, mais rápido é seu processamento. No entanto, esta correlação não foi mostrada nos transcritos de *H. salinarum* NRC-1, embora a estabilidade diferencial dos transcritos nas unidades policistrônicas se relacione as respostas adaptativas frente a mudanças ambientais (Owolabi e Rosen,

Figura 4.29: Gráficos de MA-plot para amostras das linhagens Δ ura3 e Δ VNG_RS05855 preparadas com os protocolos A) TruSeq-Stranded (amostras: RNA total depletado de rRNA); B) TruSeq-Small (amostras: RNA total depletado de rRNA, fragmentado e selecionando fragmentos entre 150 e 600 pb após ligação dos adaptadores). Os pontos laranjas são os genes com expressão diferencial para $|M| > 1$ e $A > 5$, os verdes os que têm a expressão diminuída e os pontos vermelhos expressão aumentada segundo as análises do DESeq2.

1990; Evguenieva-Hackenberg e Klug, 2009). Possivelmente, a relação do produto proteico do gene VNG_RS05855 com os transcritos policistrônicos possam sugerir a sua participação na adaptação às perturbações ambientais pelo processamento diferencial de seus membros.

Pelos gráficos do MA-plot, também distinguimos que os genes que codificam para os tRNAs tem expressão diferencial entre as linhagens. A maior quantidade deles é identificada após a fragmentação e a seleção por tamanho permite enriquecer a cobertura de reads nos transcritos menores. A presença destes genes nestas análises nos sugere que o produto proteico do gene VNG_RS05855 pode relacionar-se indiretamente ao processamento dos tRNAs, ainda que outra a proteína seja responsável pela maturação destes transcritos.

Apesar da lista de genes diferencialmente expressos entre as linhagens nocaute e controle não serem diretamente comparáveis pois foram obtidas por procedimentos diferentes (MA-plot e DESeq2), procuramos descobrir se poderiam estar relacionados a processos biológicos similares. Com esse objetivo realizamos as análises para determinar os 20 processos biológicos enriquecidos segundo os termos GO nos genes com expressão diferencial obtidos a partir dos gráficos de MA-plot.

Figura 4.30: Tamanho dos genes com expressão diferencial na linhagem Δ VNG_RS05855 segundo protocolos A) TruSeq-Small com amostras do RNA total selecionando fragmentos entre 150 a 600 pb ; B) TruSeq-Stranded com amostras do RNA total depletado de rRNA; C) TruSeq-Small com amostras do RNA total depletado de rRNA, fragmentado e selecionando fragmentos entre 150 e 600 pb.

Os principais processos relacionados aos genes com expressão diferencial entre as linhagens preparadas com o protocolo TruSeq Stranded estão vinculados ao metabolismo da glutamina e de metabólitos secundários como peptídeos não ribossomais e sideróforos (Tabela 4.7).

Tabela 4.7: Termos *GO* enriquecidos nos genes com expressão diferencial na linhagem Δ VNG_RS05855 segundo análises MA-plot nas amostras do RNA total depletado de rRNA preparadas com o protocolo TruSeq-Stranded. Além da descrição de genes com expressão diferencial que não codificam proteínas.

Termos GO	Descrição	Anotados	Significantes	Esperado	p-value
GO:0009084	glutamine family amino acid biosynthetic	8	6	1,06	1,1E-04
GO:0019748	secondary metabolic process	8	6	1,06	1,1E-04
GO:0009237	siderophore metabolic process	7	5	0,92	6,3E-04
GO:0019184	nonribosomal peptide biosynthetic process	7	5	0,92	6,3E-04
GO:0019290	siderophore biosynthetic process	7	5	0,92	6,3E-04
GO:0044550	secondary metabolite biosynthetic process	7	5	0,92	6,3E-04
GO:0009064	glutamine family amino acid metabolic.	35	12	4,62	9,9E-04
GO:0006526	arginine biosynthetic process	5	4	0,66	1,3E-03
GO:0006538	glutamate catabolic process	3	3	0,40	2,3E-03
GO:0019289	rhizobactin 1021 biosynthetic process	3	3	0,40	2,3E-03
GO:0019540	siderophore biosynthetic process	3	3	0,40	2,3E-03
GO:0019553	glutamate catabolic process	3	3	0,40	2,3E-03
GO:0019665	anaerobic amino acid catabolic process	3	3	0,40	2,3E-03
GO:0019670	anaerobic glutamate catabolic process	3	3	0,40	2,3E-03
GO:0046494	rhizobactin 1021 metabolic process	3	3	0,40	2,3E-03
GO:1902644	tertiary alcohol metabolic process	3	3	0,40	2,3E-03
GO:1902645	tertiary alcohol biosynthetic process	3	3	0,40	2,3E-03
GO:0006259	DNA metabolic process	116	26	15,31	2,7E-03
GO:0046165	alcohol biosynthetic process	6	4	0,79	3,5E-03
GO:0006525	arginine metabolic process	10	5	1,32	5,4E-03
-	tRNA	47	9	-	-
-	16S ribosomal RNA	1	1	-	-

Na preparação das amostras com RNA fragmentado utilizando protocolo TruSeq-Small, os genes com expressão diferencial entre as linhagens estão relacionados fundamentalmente aos processos de transporte de elétrons da respiração celular (Tabela 4.8).

Tabela 4.8: Termos *GO* enriquecidos nos genes com expressão diferencial na linhagem Δ VNG_RS05855 segundo análises MA-plot nas amostras do RNA total depletado de rRNA e fragmentado preparadas com o protocolo TruSeq-Small. Além da descrição de genes com expressão diferencial que não codificam proteínas

Termos GO	Descrição	Anotados	Significantes	Esperado	p-value
GO:0022904	respiratory electron transport chain	14	4	0,45	7,40E-04
GO:0045333	cellular respiration	28	5	0,91	1,53E-03
GO:0022900	electron transport chain	18	4	0,58	2,05E-03
GO:0015980	energy derivation by oxidation of organi...	32	5	1,03	2,86E-03
GO:0006026	aminoglycan catabolic process	3	2	0,10	2,99E-03
GO:0006030	chitin metabolic process	3	2	0,10	2,99E-03
GO:0006032	chitin catabolic process	3	2	0,10	2,99E-03
GO:0019646	aerobic electron transport chain	3	2	0,10	2,99E-03
GO:0046348	amino sugar catabolic process	3	2	0,10	2,99E-03
GO:1901071	glucosamine-containing compound metaboli...	3	2	0,10	2,99E-03
GO:1901072	glucosamine-containing compound cataboli...	3	2	0,10	2,99E-03
GO:0006525	arginine metabolic process	10	3	0,32	3,20E-03
GO:0006022	aminoglycan metabolic process	4	2	0,13	5,87E-03
GO:0006040	amino sugar metabolic process	5	2	0,16	9,58E-03
GO:0006526	arginine biosynthetic process	5	2	0,16	9,58E-03
GO:0009060	aerobic respiration	16	3	0,52	1,31E-02
GO:0006091	generation of precursor metabolites and ...	46	5	1,49	1,40E-02
GO:1901136	carbohydrate derivative catabolic proces...	6	2	0,19	1,41E-02
GO:0009237	siderophore metabolic process	7	2	0,23	1,93E-02
GO:0019184	nonribosomal peptide biosynthetic proces...	7	2	0,23	1,93E-02
-	tRNA	47	35	-	-
-	16S_ribosomal_RNA	1	1	-	-
-	23S_ribosomal_RNA	1	1	-	-

Apesar dos principais processos não estarem relacionados, distinguimos alguns que estão representados tanto nas amostras preparadas com o protocolo TruSeq Stranded como com o TruSeq Small: os relacionados ao metabolismo da arginina, glucosamina e a respiração. No entanto, nenhum deles está representado nos processos enriquecidos a partir dos genes que tem expressão diferencial segundo as análises do DESeq2.

4.4.2.2 Estabilidade de unidades policistrônicas podem relacionar-se com a ausência do gene VNG_RS05855

O processamento diferencial do mRNA policistrônico poderia justificar a presença da expressão diferencial de longos transcritos em amostras sem fragmentar e nas que se selecionaram fragmentos menores que 600 pb para sequenciar. O processamento diferencial do mRNA de um operon é estabelecido como um fenômeno evolutivamente conservado, caracterizado pelo papel-chave das RNases e que funciona como mecanismo de regulação dos níveis de expressão de diversos genes (Owolabi e Rosen, 1990; Naureckiene e Uhlin, 2003; Kaper e B., 2008; Dar e

Sorek, 2018).

Com estes antecedentes identificamos dentre os genes diferencialmente expressos quais estão organizados em operons e poderiam ter o processamento de seu mRNA afetado pela ausência da suposta RNase codificada pelo gene VNG_RS05855. Para isso, utilizamos a classificação feita por Price e col (Price et al., 2005) que identificou, in silico, os genes de *H. salinarum* NRC-1 organizados em operons.

Obtivemos que 36% dos genes com expressão diferencial entre as linhagens nocaute e controle, estimados pelo DESeq2 são parte de operons (Tabela 4.9). Entretanto, 34% dos genes com expressão diferencial segundo os gráficos de MA-plot, a partir das amostras preparadas com o protocolo TruSeq Stranded, também os são (Tabela 4.10). Já dos genes identificados com as mesmas análises de expressão diferencial, porém de amostras fragmentadas e preparadas com protocolo TruSeq Small, 32% são membros de operons (Tabela 4.11).

Observamos que independentemente dos protocolos utilizados para o preparo das bibliotecas serem diferentes, assim como, os métodos para estimar a expressão diferencial de genes entre as linhagens, a maioria dos membros dos operons tem a expressão aumentada na linhagem nocaute em relação à linhagem controle. Este resultado sugere que a maioria destes transcritos podem ser mais estáveis na ausência do produto proteico do gene VNG_RS05855. A ação coordenada de enzimas endo- e exo- ribonucleases para garantir o processamento diferencial dos transcritos contidos num mesmo operon é um fato caracterizado, além que sofre a influência da estrutura dos mRNA (Dar e Sorek, 2018). Os resultados sugerem que VNG_RS05855 possa exercer alguma destas atividades.

Outro aspecto relevante é que independentemente das diferenças experimentais e de tratamentos dos dados do sequenciamento, observamos alterações em operons envolvidos na produção de vesículas de gás, no transporte e na cadeia de transporte de elétrons, sugerindo que os processamentos realizados por VNG_RS05855 poderiam ser nesses genes.

Tabela 4.9: Genes com expressão diferencial, pela análise do DESeq2 nas bibliotecas preparadas com o protocolo TruSeq-Small, possíveis membros de operons. **Size:** comprimento em pb dos genes. **Member:** gene adjacente na mesma cadeia previsto no mesmo operon. **pOp:** probabilidade estimada de que o par esteja no mesmo operon (valores próximos a 1 ou 0 são previsões confiantes de estar no mesmo operon ou não, enquanto valores próximos a 0.5 são previsões de baixa confiança). **Sep:** distância entre os dois genes, em pares de bases. **MOGScore:** conservação, variando de 0 (não conservado) a 1 (100% conservado). **ExprSim:** correlação de padrões de expressão. **Function:** função do gene com expressão diferencial (Diff. Expres.)

Genes Diff. Expres.	Size(pb)	Member	pOp	Sep	MOGScore	ExprSim	Function
VNG_RS01160(up)	782	VNG_RS01165	0,86	38	0,40	0,82	MBL-fold-hydrolase
VNG_RS01165(up)	269	VNG_RS01160	0,86	38	0,40	0,82	MarR-family-transcriptional-regulator
VNG_RS01270(up)	611	VNG_RS01275	0,80	56	0,39	0,84	MarR-family-transcriptional-regulator
VNG_RS02055(up)	1214	VNG_RS02050	1,00	2	0,02	0,85	Fe-S-cluster-assembly-protein-SufD
VNG_RS02925(up)	1982	VNG_RS02930	0,95	-13	0,76	0,85	AbrB-family-transcriptional-regulator
VNG_RS02930(up)	1328	VNG_RS02935	0,97	-9	0,68	0,93	hypothetical-protein
VNG_RS02930(up)	1328	VNG_RS02925	0,95	-13	0,76	0,85	hypothetical-protein
VNG_RS02935(up)	2213	VNG_RS02930	0,97	-9	0,68	0,93	serine-protein-kinase
VNG_RS03175(up)	2450	VNG_RS03180	0,95	-3	0,00	0,84	methyl-accepting-chemotaxis-protein
VNG_RS06995(up)	515	VNG_RS07000	0,95	-3	0,00	0,78	Hsp20/alpha-crystallin-family-protein
VNG_RS07000(up)	821	VNG_RS06995	0,95	-3	0,00	0,78	hypothetical-protein
VNG_RS07475(up)	677	VNG_RS07470	0,99	1	0,04	0,70	TrkA-family-potassium-(up)take-protein
VNG_RS07785(up)	353	VNG_RS07780	0,98	4	0,26	0,88	DNA-binding-protein
VNG_RS09480(up)	1295	VNG_RS09475	0,78	70	0,23	0,75	O-acetylhomoserine-aminocarboxypropyltransferase
VNG_RS09905(up)	1043	VNG_RS09910	0,99	7	0,58	0,89	ABC-transporter-permease
VNG_RS09905(up)	1043	VNG_RS09900	0,92	58	0,64	0,81	ABC-transporter-permease
VNG_RS09910(up)	947	VNG_RS09905	0,99	7	0,58	0,89	ABC-transporter-permease
VNG_RS09910(up)	947	VNG_RS09915	0,99	2	0,50	0,87	ABC-transporter-permease
VNG_RS10585(up)	845	VNG_RS10580	0,89	0	0,00	NA	protein-gvpL
VNG_RS10585(up)	845	VNG_RS10590	0,99	-3	0,67	NA	protein-gvpL
VNG_RS10600(up)	548	VNG_RS10605	0,98	0	0,29	NA	protein-gvpH-1
VNG_RS10610(up)	641	VNG_RS10605	0,96	4	0,43	NA	protein-gvpF
VNG_RS10630(up)	1148	VNG_RS10635	0,97	-3	0,00	NA	gas-vesicle-protein-GvpC
VNG_RS10640(up)	359	VNG_RS10635	0,91	25	0,19	NA	protein-gvpO-1
VNG_RS10735(up)	995	VNG_RS10730	0,99	-7	1,00	NA	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS12280(up)	845	VNG_RS12285	0,99	-3	0,67	NA	protein-gvpL
VNG_RS12280(up)	845	VNG_RS12275	0,89	0	0,00	NA	protein-gvpL
VNG_RS12305(up)	641	VNG_RS12300	0,96	4	0,43	NA	protein-gvpF
VNG_RS12325(up)	1148	VNG_RS12330	0,97	-3	0,00	NA	gas-vesicle-protein-GvpC
VNG_RS12335(up)	359	VNG_RS12330	0,91	25	0,19	NA	protein-gvpO-1
VNG_RS12430(up)	995	VNG_RS12425	0,98	-7	1,00	NA	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS13065(up)	995	VNG_RS13070	0,99	-7	1,00	NA	cytochrome-d-ubiquinol-oxidase-subunit-II
VNG_RS04530(down)	485	VNG_RS04525	0,99	-3	0,05	0,89	nucleoside-diphosphate-kinase
VNG_RS06830(down)	713	VNG_RS06835	0,98	2	0,00	0,79	sensory-rhodopsin-2
VNG_RS09710(down)	1115	VNG_RS09705	1,00	0	0,20	0,93	phosphate-ABC-transporter-permease-subunit-PstC
VNG_RS11030(down)	1985	VNG_RS11035	0,98	-3	0,05	NA	ORF-H1161
VNG_RS11035(down)	2939	VNG_RS11030	0,98	-3	0,05	NA	ORF-H1186
VNG_RS11035(down)	2939	VNG_RS11040	0,99	-3	0,03	NA	ORF-H1186
VNG_RS11195(down)	818	VNG_RS11190	0,88	0	0,00	NA	universal-stress-protein
VNG_RS12730(down)	2939	VNG_RS12735	0,99	-3	0,03	NA	ORF-H1186
VNG_RS12735(down)	545	VNG_RS12730	0,99	-3	0,03	NA	transcription-factor
VNG_RS12855(down)	1934	VNG_RS12860	0,99	1	0,57	NA	arsenic-transporting-ATPase
VNG_RS12900(down)	512	VNG_RS12895	0,99	1	0,88	NA	ORF-H1517

Tabela 4.10: Seleção de 43 genes de um total de 139 com expressão diferencial (Anexo 7.2, restantes genes), segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Stranded, identificados como possíveis membros de operons. **Size:** comprimento em pb dos genes. **Member:** gene adjacente na mesma cadeia previsto no mesmo operon. **pOp:** probabilidade estimada de que o par esteja no mesmo operon (valores próximos a 1 ou 0 são previsões confiantes de estar no mesmo operon ou não, enquanto valores próximos a 0.5 são previsões de baixa confiança). **Sep:** distância entre os dois genes, em pares de bases. **MOGScore:** conservação, variando de 0 (não conservado) a 1 (100% conservado). **ExprSim:** correlação de padrões de expressão. **Function:** função do gene com expressão diferencial (Diff. Expres.)

Genes Diff. Expres.	Size(pb)	Member	pOp	Sep	MOGScore	ExprSim	Function
VNG_RS02230(up)	326	VNG_RS02235	0,99	-3	0,66	NA	Na+/H+_antiporter_subunit_G
VNG_RS02230(up)	326	VNG_RS02225	0,98	-3	0,09	NA	Na+/H+_antiporter_subunit_G
VNG_RS02235(up)	275	VNG_RS02230	0,99	-3	0,66	NA	cation:proton_antiporter
VNG_RS02235(up)	275	VNG_RS02240	1,00	-3	0,13	0,914	cation:proton_antiporter
VNG_RS03255(up)	788	VNG_RS03250	1,00	-3	1,00	0,872	4Fe-4S_ferredoxin
VNG_RS03255(up)	788	VNG_RS03260	1,00	0	1,00	0,871	4Fe-4S_ferredoxin
VNG_RS03385(up)	1082	VNG_RS03380	0,99	0	0,09	0,783	asparagine_synthase
VNG_RS03475(up)	2474	VNG_RS03470	1,00	2	0,31	0,692	DNA_gyrase_subunit_A
VNG_RS06440(up)	719	VNG_RS06435	1,00	0	0,50	0,788	rhodopsin
VNG_RS06620(up)	527	VNG_RS06625	0,99	-3	0,00	0,964	50S_ribosomal_protein_L5
VNG_RS06620(up)	527	VNG_RS06615	1,00	-3	0,07	0,962	50S_ribosomal_protein_L5
VNG_RS06625(up)	158	VNG_RS06620	0,99	-3	0,00	0,964	30S_ribosomal_protein_S14
VNG_RS06625(up)	158	VNG_RS06630	0,99	3	0,00	0,954	30S_ribosomal_protein_S14
VNG_RS09235(up)	980	VNG_RS09240	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS09235(up)	980	VNG_RS09230	0,98	27	1,00	0,895	peptide_ABC_transporter_permease
VNG_RS09240(up)	1160	VNG_RS09235	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS10730(up)	1448	VNG_RS10735	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_I
VNG_RS10735(up)	995	VNG_RS10730	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS12125(up)	995	VNG_RS12130	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS12430(up)	995	VNG_RS12425	0,98	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS13065(up)	995	VNG_RS13070	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS11355(up)	359	VNG_RS11360	0,85	2	0,00	NA	protein_gvpK
VNG_RS11360(up)	296	VNG_RS11365	0,98	-3	0,50	NA	gas_vesicle_protein_GvpJ
VNG_RS11360(up)	296	VNG_RS11355	0,85	2	0,00	NA	gas_vesicle_protein_GvpJ
VNG_RS11400(up)	857	VNG_RS11405	0,97	-3	0,00	NA	gas_vesicle_protein_GvpC
VNG_RS11405(up)	1034	VNG_RS11400	0,97	-3	0,00	NA	ATPase_AAA
VNG_RS11405(up)	1034	VNG_RS11410	0,92	23	0,19	NA	ATPase_AAA
VNG_RS11185(up)	2159	VNG_RS11190	0,99	0	0,00	NA	potassium-transporting_ATPase_subunit_B
VNG_RS11185(up)	2159	VNG_RS11180	0,98	2	0,00	NA	potassium-transporting_ATPase_subunit_B
VNG_RS11190(up)	650	VNG_RS11195	0,88	0	0,00	NA	ATPase
VNG_RS11190(up)	650	VNG_RS11185	0,99	0	0,00	NA	ATPase
VNG_RS11285(up)	1574	VNG_RS11290	0,98	-3	0,28	NA	aspartate_aminotransferase_family_protein
VNG_RS11290(up)	1823	VNG_RS11285	0,98	-3	0,28	NA	iron_transporter
VNG_RS11290(up)	1823	VNG_RS11295	0,98	-3	0,28	NA	iron_transporter
VNG_RS11300(up)	1433	VNG_RS11305	0,78	26	0,06	NA	lysine_6-monooxygenase
VNG_RS11300(up)	1433	VNG_RS11295	0,97	-3	0,29	NA	lysine_6-monooxygenase
VNG_RS11305(up)	1826	VNG_RS11300	0,78	26	0,06	NA	iron_transporter
VNG_RS00415(down)	1499	VNG_RS00420	0,91	-7	0,00	0,301	RmeM
VNG_RS01040(down)	599	VNG_RS01045	0,97	-7	0,82	0,596	ribonuclease_H
VNG_RS01795(down)	284	VNG_RS01800	0,95	2	0,00	0,754	transcriptional_regulator
VNG_RS02370(down)	1085	VNG_RS02365	0,77	119	0,06	0,884	glycosyl_transferase_family_1
VNG_RS05025(down)	1613	VNG_RS05020	1,00	0	1,00	0,852	potassium_transporter_Trk
VNG_RS07810(down)	725	VNG_RS07815	0,96	5	0,01	0,86	uridylylate_kinase

Tabela 4.11: Genes com expressão diferencial, segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Small fragmentado, identificados como possíveis membros de operons. **Size:** comprimento em pb dos genes. **Member:** gene adjacente na mesma cadeia previsto no mesmo operon. **pOp:** probabilidade estimada de que o par esteja no mesmo operon (valores próximos a 1 ou 0 são previsões confiáveis de estar no mesmo operon ou não, enquanto valores próximos a 0.5 são previsões de baixa confiança). **Sep:** distância entre os dois genes, em pares de bases. **MOGScore:** conservação, variando de 0 (não conservado) a 1 (100% conservado). **ExprSim:** correlação de padrões de expressão. **Function:** função do gene com expressão diferencial (Diff. Expres.)

Genes Diff. Expres.	Size(pb)	Member	pOp	Sep	MOGScore	ExprSim	Function
VNG_RS02205(up)	1619	VNG_RS02200	0,99	-3	0,12	0,372	cation:proton_antipporter
VNG_RS02205(up)	1619	VNG_RS02210	0,99	-3	0,22	0,331	cation:proton_antipporter
VNG_RS02210(up)	1481	VNG_RS02205	0,99	-3	0,22	0,331	cation:proton_antipporter
VNG_RS02925(up)	1982	VNG_RS02930	0,95	-13	0,76	0,85	AbrB_family_transcriptional_regulator
VNG_RS02930(up)	1328	VNG_RS02925	0,95	-13	0,76	0,85	hypothetical_protein
VNG_RS02930(up)	1328	VNG_RS02935	0,97	-9	0,68	0,926	hypothetical_protein
VNG_RS02935(up)	2213	VNG_RS02930	0,97	-9	0,68	0,926	serine_protein_kinase
VNG_RS03200(up)	1718	VNG_RS03205	0,98	-3	0,00	0,565	chitinase
VNG_RS03475(up)	2474	VNG_RS03470	1,00	2	0,31	0,692	DNA_gyrase_subunit_A
VNG_RS04465(up)	749	VNG_RS04460	0,98	-3	0,01	0,957	30S_ribosomal_protein_S2
VNG_RS06070(up)	698	VNG_RS06075	0,81	27	0,50	NA	hypothetical_protein
VNG_RS06070(up)	698	VNG_RS06065	0,99	-3	0,90	0,957	hypothetical_protein
VNG_RS06080(up)	371	VNG_RS06075	0,98	-3	0,70	NA	hypothetical_protein
VNG_RS09230(up)	1961	VNG_RS09235	0,98	27	1,00	0,895	peptide-binding_protein
VNG_RS09235(up)	980	VNG_RS09240	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS09235(up)	980	VNG_RS09230	0,98	27	1,00	0,895	peptide_ABC_transporter_permease
VNG_RS09240(up)	1160	VNG_RS09235	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS09525(up)	488	VNG_RS09530	0,82	16	0,00	0,943	hypothetical_protein
VNG_RS09530(up)	1457	VNG_RS09535	0,99	-3	0,01	0,896	argininosuccinate_lyase
VNG_RS09530(up)	1457	VNG_RS09525	0,82	16	0,00	0,943	argininosuccinate_lyase
VNG_RS09535(up)	1190	VNG_RS09530	0,99	-3	0,01	0,896	argininosuccinate_synthase
VNG_RS09775(up)	1703	VNG_RS09770	0,99	0	0,05	0,134	phenylalanine-tRNA_ligase_subunit_beta
VNG_RS10435(up)	1193	VNG_RS10430	0,98	0	0,01	0,62	DNA-directed_RNA_polymerase_subunit_A
VNG_RS10595(up)	344	VNG_RS10590	0,99	2	0,75	NA	gas_vesicle_protein_GvpJ
VNG_RS11300(up)	1433	VNG_RS11295	0,97	-3	0,29	NA	lysine_6-monooxygenase
VNG_RS11300(up)	1433	VNG_RS11305	0,78	26	0,06	NA	lysine_6-monooxygenase
VNG_RS11305(up)	1826	VNG_RS11300	0,78	26	0,06	NA	iron_transporter
VNG_RS11405(up)	1034	VNG_RS11400	0,97	-3	0,00	NA	ATPase_AAA
VNG_RS11405(up)	1034	VNG_RS11410	0,92	23	0,19	NA	ATPase_AAA
VNG_RS11610(up)	479	VNG_RS11605	0,96	-22	0,20	NA	aspartate_carbamoyltransferase_regulatory_subunit
VNG_RS12130(up)	1448	VNG_RS12125	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_I
VNG_RS13065(up)	995	VNG_RS13070	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS01645(down)	392	VNG_RS01650	0,99	0	0,31	0,873	NagC_family_transcriptional_regulator

4.5 Análises em larga escala dos sítios que marcam os inícios de transcrição e de processamentos do RNA (dRNAseq)

O mapeamento de possível sítios que indicam inícios da transcrição ou posições de processamento pós-transcricional requer a capacidade de discriminar entre transcritos primários e

processados (Sharma e Vogel, 2014; Xu et al., 2015). Em organismos procariotos, os transcritos primários ou recém-sintetizados se caracterizam por ter sua extremidade 5' tri-fosfatadas, no entanto os transcritos processados têm esse extremo mono-fosfatado (Qi et al., 2017).

A utilização do differential RNA-seq (dRNA-seq) têm permitido identificar os sítios que marcam os inícios de transcrição (TSS-transcription start site) de forma global numa variedade organismos procariotos (Sharma et al., 2010; Sharma e Vogel, 2014; Ten-Caten et al., 2018). Basicamente, pela comparação dos sequenciamentos de amostras enriquecidas com RNA primário (TEX+) com o controle (TEX-) que contém o conjunto total dos RNAs primários e processados. A comparação entre os produtos de sequenciamento pode ser feita pelo software TSSAR stand-alone que permite a detecção automatizada dos sítios que marcam os inícios da transcrição (TEX + > TEX -) ((Amman et al., 2014; Ten-Caten et al., 2018); Amr et al., em preparação).

A aplicabilidade do dRNA-seq tem sido estendida na determinação de sítios de processamento do transcrito (TPS- transcript processing sites) ((Xu et al., 2015; Qi et al., 2017); Amr et al., em preparação). Com software TSSAR stand-alone também é possível determinar automatizadamente os TPS, porém invertendo forçadamente os rótulos de bibliotecas enriquecidas e não enriquecidas (TEX - > TEX +). Ao trabalhar com os mesmos resultados do sequenciamento não vai existir sobreposição entre sítios que marquem o início da transcrição e os que marquem o processamento dos transcritos. Ambos serão extremidades 5', trifosfatados para o início da transcrição e monofosfatados para os sítios de processamento. Entre estes quadros não há uma diferença substancial em termos de modelos estatísticos (Amr et al., em preparação).

4.5.1 Identificação dos sítios de início da transcrição (TSS)

Nosso primeiro passo foi verificar se com dados do sequenciamento dRNA-seq conseguimos reproduzir as características dos sítios que marcam os inícios da transcrição em ambas linhagens. Com apoio da ferramenta TSSAR stand-alone definimos os inícios de transcrição (TSS) fixando como valor de corte o p-valor de 0.001, o número mínimo de 10 reads por posição e o agrupamento de inícios de transcrição identificados a menos de 10 pares de base de distancia (Ten-Caten et al., 2018). Trabalhamos com os TSS primários, posicionado 250 nt upstream do gene anotado; os internos, posicionado dentro da região codificadora do gene; os

antisense, que está na cadeia oposta do gene anotado (Sharma e Vogel, 2014; Amman et al., 2014).

No total foram identificados 1607 possíveis TSS na linhagem controle e 448 na linhagem nocaute (Figura 4.31). A diferença na quantidade de TSS entre as linhagens pode ter sua origem no baixo número de reads na biblioteca TEX+ da linhagem nocaute (Tabela 4.3). No entanto, nota-se que apesar de ser baixo número de reads é possível identificar as características das regiões de início de transcrição e promotoras nos TSS identificados na linhagem nocaute (Figura 4.31 B). Observamos a alta frequência de purinas na posição de início de transcrição (0), pirimidinas na posição -1 e a presença de nucleotídeos que compõem o start códon ATG. Nas regiões promotoras conseguimos identificar a TATA-box, centrada entre os 24-28nt upstream ao TSS, e adjacentes a ela regiões ricas em purinas que correspondem aos elementos BRE (Gregor e Pfeifer, 2005). Estes resultados reproduzem as características dos TSS, apesar da deficiência do número de reads sequenciados na linhagem nocaute, assim, é possível considerar que também com esses dados seja possível identificar os prováveis sítios de processamento pós-transcricional nesta linhagem.

Nota-se que para estimar as sequências conservadas utilizamos a distribuição de nucleotídeos de *Thermus thermophilus*. Este organismo tem uma distribuição do conteúdo GC similar a *H. salinarum* NRC-1 (cromossomo de *T. thermophilus* ~70% do GC por ~68% de *H. salinarum* NRC-1 ; elementos plasmidiais de *T. thermophilus* ~69% GC por ~ 59% de *H. salinarum* NRC-1), o que permite realizar uma ajuste na relação da entropia (bits) com a aparição de monômeros na sequência, tendo a aparição de monômeros raros um maior conteúdo de informação do que monômeros relativamente comuns (Crooks et al., 2004).

Figura 4.31: Análises dos resultados do dRNA-seq com a ferramenta TSSAR stand-alone (Amman et al., 2014) para determinar inícios da transcrição (TSS) nas linhagens A) Δ ura3 e B) Δ VNG_RS05855. Para cada linhagem se apresentam em diagrama de Venn as quantidades de TSS primários, internos, antisenses e das interseções entre estas classificações. Mostramos as representações da conservação das sequências de nucleotídeos na vizinhança do total de TSS e dos TSS primários, analisando 5 posições *downstream* e 40 posições *upstream* ao TSS, ao que definimos como a posição 0. A representação foi feita com o pacote de R “RWebLogo” (Wagih, 2016) considerando a distribuição de nucleotídeos de “*T. thermophilus*” por ser similar a *H. salinarum* NRC-1.

4.5.2 Identificação dos sítios processamento do transcrito (TPS)

Para a identificação automatizada dos TPS (transcriptional processing sites) continuamos a trabalhar com a ferramenta TSSAR stand-alone (Amman et al., 2014) e fixamos como corte o p-valor de 0.001, número mínimo de 10 reads por posição e o agrupamento de sítios de processamento identificados a menos de 10 pares de base de distância. Identificamos os TPS primários, posicionado 250 nt upstream do gene anotado; os internos, posicionado dentro da região codificadora do gene; os antisense, que está na cadeia oposta do gene anotado. Analisamos a frequência de nucleotídeos na vizinhança dos TPS, 10 nucleotídeos upstream e downstream ao sítio TPS como foi sugerido por (Qi et al., 2017)(Figura 4.32).

Figura 4.32: Análises dos resultados do dRNA-seq com a ferramenta TSSAR stand-alone (Amman et al., 2014) para determinar sítios de processamento dos transcritos (TPS) nas linhagens A) Δ ura3 e B) Δ VNG_RS05855. Para cada linhagem se apresentam em diagrama de Venn as quantidades de TPS primários, internos, antisense e das interseções entre estas classificações. Mostramos as representações da conservação das sequências de nucleotídeos na vizinhança do conjunto total de sítios TPS, assim como os TPS primários, internos e antisense, analisando as 10 posições *downstream* e *upstream* ao sítio de processamento. A representação foi feita com o pacote de R “RWebLogo” (Wagih, 2016) considerando a distribuição de nucleotídeos de “*T. thermophilus*” por ser similar a *H. salinarum* NRC-1.

Observamos que a linhagem nocaute tem menos TPS que a linhagem controle, o que poderia ser um resultado esperado devido à ausência de uma proteína com suposta função ribonucleolítica. Além disso, os TPS primários são os mais abundantes, contrário ao que obser-

vamos na linhagem controle onde os TPS internos estão em maior quantidade. Esta diferença pode sugerir que o produto proteico do gene VNG_RS05855 apoie o processamento interno das moléculas de RNA. Ainda assim, nós não descartamos que este efeito tenha sua origem nas diferenças de reads sequenciados entre as linhagens obtidos do processamento do dRNA-seq.

Ao estudar a conservação das sequências no entorno dos possíveis TPS nenhum sinal específico foi identificado, nem na linhagem controle nem na linhagem nocaute (Figura 4.32). O resultado é razoável visto que estamos avaliando a atividade do conjunto total de RNases, sendo muito provável que mais de um processamento esteja em ação ao mesmo tempo.

4.5.3 Possíveis TPS do produto proteico do gene VNG_RS05855

Identificados os TPS nas linhagens controle e nocaute, determinamos quais deles poderiam estar associados a VNG_RS05855. Para isso, comparamos as posições dos TPS entre as linhagens e distinguimos os presentes unicamente na linhagem controle, sugerindo que a sua presença seja pela atividade do produto proteico do gene VNG_RS05855, ausente na linhagem nocaute.

Com esta abordagem, conseguimos identificar 591 possíveis TPS, produto da atividade da proteína codificada no gene VNG_RS05855, que classificamos como primários, internos e antisense (Figura 4.33). Estes TPS também estão distribuídos no conjunto de replicons de *H. salinarum* NRC-1 desigualmente (Figura 4.33 B), sendo no cromossomo onde se apresentam em menor densidade (Figura 4.33 C).

Os TPS internos são mais abundantes, seguidos pelos primários e por último os antisense, que são localizados na cadeia oposta do gene anotado (Figura 4.33 A). A presença de TPS internos associados à atividade do produto proteico do gene VNG_RS05855 apoia a sugestão da sua possível função no processamento interno dos RNA. Por outro lado, TPS primários permitem indicar que este produto proteico poderia participar do processamento de sequências próximas ao códon de início da tradução ou upstream nos sítios de união do ribossomo e relacionar-se com a função reguladora do início da tradução. Esta atividade poderia também ser realizada por meio de TPS antisense, regulando a produção de pequenos RNA antisense (asRNA) que além de regular as taxas de tradução do RNA, poderiam influenciar a estabilidade dos mRNA (Brantl, 2002; Sesto et al., 2013; Almeida et al., 2019).

Figura 4.33: Sítios de processamento pós-transcricional (TPS) do produto proteico do gene VNG_RS05855. A) Diagrama de Venn com as quantidades de TPS primários, internos, anti-sense e das interseções entre estas classificações. B) Número de TPS presentes nos RNA dos replicons de *H. salinarum*, a representação não está em escala respeito aos tamanhos dos replicons. C) Densidade de TPS em cada replicon estimada como a relação entre o seu número e o tamanho de cada replicon. D) Representações da conservação das sequências de nucleotídeos na vizinhança dos sítios TPS primário, internos e antisense, analisando as 10 posições downstream e upstream ao sítio TPS. A representação foi feita com o pacote de R “RWebLogo”(Wagih, 2016) considerando a distribuição de nucleotídeos de “*T. thermophilus*” por ser similar a *H. salinarum* NRC-1

Avaliamos se existem sequências conservadas nos TPS que sirvam de sinais para o reconhecimento do produto proteico do gene VNG_RS05855. Porém, com nossas análises não achamos sequências conservadas (Figura 4.33 D).

Conseguimos detectar transcritos ou segmentos de transcritos que foram clivados de precursores mais longos, porém não podemos elucidar o mecanismo dessa quebra molecular. É impossível saber se os transcritos detectados são resultado da digestão exoribonucleolítica 5'-3', da paralisação desta atividade ou da clivagem endoribonucleolítica; em todos os casos a extremidade 5' é o que se observa. Para o dRNA-seq, que permite o enriquecimento de reads na extremidade 5', não é esperado o enriquecimento desta extremidade pela ação exoribonucleolítica 3'-5', pois esta atividade mantém o produto da digestão na extremidade 5' inalterável.

4.5.3.1 TPS nos genes com expressão diferencial

A partir do grupo de genes com expressão diferencial entre as linhagens, estimados pelo DESeq2, localizamos os possíveis TPS na sua vizinhança. O objetivo destas análises é, mediante o posicionamento dos TPS, contextualizar a possível atividade do produto proteico do gene VNG_RS05855 nas variações de expressão desses genes. Este fato é possível porque as bibliotecas utilizadas para a estudo dos genes com expressão diferencial, assim como as preparadas para estudar o enriquecimento de RNA primários não tiveram fragmentação de amostras.

No operon dpp localizamos três TPS, dois upstream aos genes com expressão diferencial e um terceiro localizado interno ao gene VNG_RS09905, próximo ao sítio que marca o início da transcrição (Figura 4.34 A). Previamente, tínhamos sugerido, segundo as diferenças nos sinais de expressão gênica, os aumentos na estabilidade da extremidade 3' deste operon na ausência da possível RNase VNG_RS05855. A localização dos TPS na extremidade 5' desta unidade policistônica é mais um argumento que a região possivelmente seja processada. As análises dos TPS não permitem identificar se o processamento é consequência da parada da sua possível atividade exo- ou endo- ribonucleolítica, porém, é possível distinguir que, a partir deles, e na direção da transcrição, a expressão diferencial dos genes entre as linhagens, indicando a ausência de processamento na linhagem nocaute. Assim, no mesmo operon seus membros podem ter um processamento diferencial, talvez pela redundância nas enzimas que os processam.

Figura 4.34: Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 determinados segundo análises dos dados dRNA-seq e indicados em triângulos pretos. Triângulos cinzas representam sítios de processamento determinados “*in house*”. Mostramos a cobertura de reads das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*), e das amostras do RNA total fragmentado após da depleção do rRNA com os protocolos TruSeq mRNA Stranded (**) e TruSeq-Small (***) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019). Em negrito os genes com expressão diferencial. Representamos os transportadores ABC que participam A) no deslocamento de dipeptídeos (*dpp*), B) de fosfato (*pst*).

Identificamos TPS em outro operon, também relacionado ao transporte, mas desta vez com expressão diminuída na linhagem nocaute respeito à linhagem controle. O operon *pst*, relacionado a captação de fosfato extracelular, tem o gene VNG_RS09710 com expressão diminuída. Nele, os TPS estão localizados internos a outros genes do operon downstream ao gene com expressão diferencial (Figura 4.34 B). Com base na função repressora conhecida das RNases, a deleção de um gene relacionado a essa função resultaria predominantemente na regulação positiva de seus genes alvos (Wurtmann et al., 2014), porém é mais difícil encontrar explicações ao efeito contrário, ou seja, genes que incrementam a sua expressão na linhagem nocaute de uma RNase. Porém, percebemos que a localização dos TPS é diferente a localização dos TPS nos que estão com expressão aumentada o que poderia, de alguma maneira, justificar as diferenças observadas na expressão dos genes.

No operon das vesículas de gás identificamos TPS não só em regiões upstream aos ge-

Figura 4.35: Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 determinados segundo análises dos dados dRNA-seq e indicados em triângulos pretos. Triângulos cinzas representam sítios de processamento determinados “in house”. Mostramos a cobertura de reads das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*), e das amostras do RNA total fragmentado após da depleção do rRNA com os protocolos TruSeq mRNA Stranded (**) e TruSeq-Small (***) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019). caixas vermelhas os resultados da interação da proteína Lsm com RNA em ensaios de RIP-seq (Zaramela, 2015). Em negrito os genes com expressão diferencial. Representamos os membros do operon das vesículas de gás.

nes com expressão aumentada na linhagem nocaute em relação à linhagem controle (genes *gvpC* e *gvpF*), também identificamos TPS dentro de suas regiões codificadoras (genes *gvpA*, *D*, *O*, *F*, *L*) (Figura 4.35). No operon de vesículas de gás os genes *gvpF*, *D*, *A* e *O* tem promotores alternativos (Pfeifer, 2012). Deles, o único que tem TPS upstream é o gene *gvpF*, o processamento pós-transcricional da região 5'-UTR poderia afetar os sítios de reconhecimento dos ribossomos nos transcritos alternativos. No entanto, para os genes *gvpA* e *D* os TPS são internos nos seus RNAs, sugerindo a eventual participação do produto proteico do gene VNG_RS05855 na degradação. Para o gene *gvpO* o TPS identificado é antisense a ele, numa região anotada com um possível RNA antisense (Almeida et al., 2019), o que supostamente aponta a relação do produto proteico do gene VNG_RS05855 com a regulação antisense também desse gene. No entanto, não foi detectado TPS antisense ao gene *gvpA*.

Analizamos os TPS no regulador transcricional da família MarR (VNG_RS01165).

Figura 4.36: Possíveis sítios de processamento do produto proteico do gene VNG_RS05855 determinados segundo análises dos dados dRNA-seq e indicados em triângulos pretos. Triângulos cinzas representam sítios de processamento determinados “*in house*”. Mostramos a cobertura de reads das bibliotecas preparadas a partir de amostras do RNA total depletado de rRNA com o protocolo TruSeq-Small (*), e das amostras do RNA total fragmentado após da depleção do rRNA com os protocolos TruSeq mRNA Stranded (**) e TruSeq-Small (***) das linhagens controle (verde) e nocaute (vermelho). A cobertura de *reads* descrevem a média das contagens das réplicas por linhagem. Caixas amarelas representam genes anotados na fita *forward* e laranjas na fita *reverse*. As caixas roxas representam RNA antisense anotados no genoma de *H. salinarum* NRC-1 (Almeida et al., 2019). Em negrita os genes com expressão diferencial. Representamos os membros A) do operon MarR e B) Halorodopsina.

Observamos que além de ser interno e localizado na extremidade 3' do transcrito, o que poderia indicar a possível função na regulação da sua estabilidade, este TPS é upstream a outro gene que, além de ter a regulação aumentada na linhagem nocaute respeito ao controle, forma o operon com ele (Figura 4.36). Este resultado contrasta com a localização dos TPS ao estudar os membros dos transportadores Dpp com expressão diferencial, no entanto é um indício da complexidade e variações da regulação pós-transcricional de genes que codificam reguladores transcricionais.

Apresentamos a cobertura de reads do gene halorodopsina que não tem a expressão transcricional afetada pela ausência de VNG_RS05855. No entanto, nele observamos que na sua região 3'-UTR a cobertura reads entre as linhagens é diferente (Figura 4.36). Este fato, que se repete em outras regiões do genoma, nos leva a propor que regiões não traduzidas dos genes contêm informação não recuperadas pelas análises de expressão diferencial. Estas informações deverão ser exploradas cuidadosamente no futuro com o objetivo de aprofundar as variações da

regulação pós-transcricional na linhagem nocaute do gene VNG_RS05855.

Nos resultados apresentados nesta seção, primeiramente, analisamos a cobertura de reads para as bibliotecas preparados com o protocolo TruSeq Small sem fragmentar. Em seguida, decidimos também inserir a cobertura de reads dos sequenciamentos das amostras obtidas dos protocolos TruSeq Stranded e TruSeq Small com fragmentação na visualização do contexto onde são inseridos os TPS. Observamos que com ou sem fragmentação das amostras, a cobertura de reads para as regiões que contêm TPS contem características similares.

4.5.3.2 Processos biológicos associados aos genes com TPS

Após de identificar os possíveis TPS, verificamos os genes que os contêm mudando em 20nt as coordenadas deles com o objetivo de estudar as regiões não traduzidas. Com este enfoque conseguimos distinguir que os TPS estão em 419 genes, cujos transcritos podem ser potenciais do VNG_RS05855. Este número de genes representam aproximadamente 14% dos genes anotados em *H. salinarum* NRC-1. Realizamos a análise de processos biológicos representados nesses genes (Tabela 4.12). Distinguimos TPS nos RNA de proteínas relacionadas à locomoção, quimiotaxia, resposta a zinco e metabolismo de purinas. Estes processos biológicos são diferentes aos que determinamos a partir das listas de genes com expressão diferencial obtidos pelas análises do DESeq2 e MA-plot.

Tabela 4.12: Termos *GO* enriquecidos nos genes com TPS do produto proteico do gene VNG_RS05855.

Termo GO	Descrição	Anotados	Significantes	Esperado	p-valor
GO:0040011	locomotion	34	13	5,97	3,0E-03
GO:0006935	chemotaxis	32	12	5,62	5,3E-03
GO:0042330	taxis	32	12	5,62	5,3E-03
GO:0010043	response to zinc ion	3	3	0,53	5,4E-03
GO:0009127	purine nucleoside biosynthesis	33	12	5,80	7,1E-03
GO:0009168	purine ribonucleoside biosynthesis	33	12	5,80	7,1E-03
GO:0022607	cellular component assembly	42	14	7,38	8,9E-03
GO:0006164	purine nucleotide biosynthetic process	38	13	6,68	9,1E-03
GO:0072522	purine-containing compound biosynthetic	38	13	6,68	9,1E-03
GO:0016043	cellular component organization	91	25	15,99	9,6E-03
GO:0009126	purine nucleoside metabolic	51	16	8,96	1,0E-02
GO:0009167	purine ribonucleoside metabolic	51	16	8,96	1,0E-02
GO:0010038	response to metal ion	6	4	1,05	1,0E-02
GO:0065009	regulation of molecular function	6	4	1,05	1,0E-02
GO:0072521	purine-containing compound metabolic	66	19	11,60	1,4E-02
GO:0008219	cell death	4	3	0,70	1,9E-02
GO:0010941	regulation of cell death	4	3	0,70	1,9E-02
GO:0030003	cellular cation homeostasis	4	3	0,70	1,9E-02
GO:0031399	regulation of protein modification	4	3	0,70	1,9E-02
GO:0043086	negative regulation of catalytic activity	4	3	0,70	1,9E-02
-	tRNA	47	7	-	-

Por outro lado, 49 dos 117 genes com expressão diferencial identificados pelo DESeq2, tem TPS associados a seus transcritos (Figura 4.33). Esses genes estão relacionados com a produção de vesículas de gás e com respostas a estímulos ambientais. 41 genes com expressão diferencial segundo os protocolos de preparo das bibliotecas (TruSeq-Stranded-28- e TruSeq-small fragmentado-13-) analisados pelo MA-plot tem TPS associados. Eles estão relacionados com tRNA e o transporte de íons. A distribuição desigual de TPS em genes que representam distintos processos biológicos sugere que o papel do produto proteico do gene VNG_RS05855 seja pela variedade, inespecífico, e pelo fato de não ter associado um fenótipo claro na sua ausência, redundante (Figura 4.37).

Figura 4.37: Número de genes com expressão diferencial (cor: magenta) determinados pelos análises do DESeq2 (TruSeq-small) e MA-plot (TruSeq Stranded- e TruSeq Small- fragmentado). Em azul a porcentagem de genes com TPS.

4.5.3.3 Estrutura secundária das sequências que contem os TPS

A estrutura do RNA, mais que a sua sequência, desempenham um papel fundamental no seu reconhecimento por proteínas, tais como fatores de regulação, chaperonas do RNA ou enzimas ribonucleolíticas (Daou-Chabo e Condon, 2009; Le Laz, Le Goaziou e Henneke, 2010; Bandyra, Wandzik e Luisi, 2018). Com base neste conhecimento dirigimos nosso esforços a estudar se existem estruturas conservadas nas regiões que têm os TPS. Para isso, fixamos a partir do TPS as sequências a serem estudadas criando uma janela de 50 nt (25nt *upstream* e *downstream*), ampliando em 15 nt a janela de estudo proposta (Qi et al., 2017).

Com o conjunto das sequências trabalhamos com os software Mfold off-line (v.3.6) que prevê as estruturas secundárias do RNA pesquisando o estado associado com a sua mínima energia livre (Minimun Free Energy – kcal/mol-) (Zuker, 2003). Neste trabalho mantivemos inalteráveis os parâmetros padrões do programa. Determinamos que a frequência dos valores de mínima energia, para o conjunto de 591 sequências nas que estão contidos os TPS, se ajustam a uma distribuição normal com média de -13.27 kcal/mol (Figura 4.38). Os valores negativos da energia livre sugerem que os TPS podem estar contidos em estruturas termodinamicamente favoráveis.

Figura 4.38: Histograma das mínimas energias livres (mfe-kcal/mol-) para o conjunto de sequências que contêm os TPS e a curva da distribuição normal dessa variável.

Para determinar se existem estruturas secundárias consensos, classificamos o conjunto de sequências contendo os TPS em sequências primárias, internas e antisense segundo prévia classificação. Cada conjunto foi trabalhado independentemente com a ferramenta RNAclust

versão 1.3. (Engelhardt et al., 2010). Esta ferramenta faz o agrupamento de moléculas de RNA baseados nos motivos de suas estruturas secundárias por meio do cálculo de todos os possíveis pares de alinhamentos, com apoio da ferramenta LocARNA (Will et al., 2012). Após do cálculo do alinhamento, RNAclust calcula a matriz de probabilidade da distribuição de estruturas secundárias apoiado no programa RNAfold do pacote Vienna (Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, 2011). Baseado no cálculo da matriz de probabilidade, RNAclust é capaz de identificar os nós relevantes da árvore formada pelo agrupamento das sequências que compartilham estruturas secundárias consensos. Para o cálculo da mínima energia livre (MFE) associada a estrutura secundária consenso de cada nó dentro da árvore do agrupamento usamos o programa RNAsoup (SpotgrOUPs in RNA cluster-tree)(Reiche, 2008) da ferramenta RNAclust, que faz esta estimativa a partir da avaliação do erro ao quadrado das energias livres mínimas das sequências individuais.

Com estas análises conseguimos identificar agrupamentos das estruturas secundárias consensos nas sequências que contêm os TPS primários, internos e antisense. De cada agrupamento selecionamos os nós com 5 e 12 sequências para estudar a conservação da estrutura. Esta seleção é arbitrária, porém responde a um critério empírico de diminuir as possibilidades de ter uma estrutura consenso estudando um número alto de sequências. Por outro lado, as estruturas secundárias propostas baseadas em um número baixo de sequências pode não abranger a total diversidade de estruturas. Nesses nós, analisamos estrutura secundária consenso e a mínima energia livre associada a ela.

4.5.3.3.1 Não há conservação nas estruturas das sequências que contem os TPS associados ao gene VNG_RS05855

Os TPS primários se localizam, geralmente, nas regiões não pareadas das estruturas conhecidas como stem-loop (Figura 4.39). Nestas mesmas regiões são preditas a maioria dos sítios de processamento das regiões 5'-UTR dos RNA de proteínas ribossomais de arqueias metanogênicas (Qi et al., 2017). Estas regiões estruturadas, geralmente anteriores aos sítios que marcam o início da tradução, reforça a ideia do papel regulador de VNG_RS05855. Contudo, não existe uma estrutura claramente definida que sirva como sinal para o processamento. De maneira similar, as estruturas dos TPS internos apresentam padrão de estruturas ainda mais he-

terogêneo (Figura 4.40), com TPS não só nas regiões não pareadas do stem-loop como, também na dupla fita que forma estas estruturas, e em regiões lineares de sequência. Esta diversidade pode ser explicada pela variabilidade das sequências codificadoras, em relação a regiões 5'-UTR que, geralmente, devem conservar a informações que marcam o início da tradução. Por outro lado, as estruturas dos RNA que contém TPS antisense não só se caracterizam pela sua diversidade (Figura 4.41), também por ter menos TPS que as categorias restantes.

Figura 4.40: Análises das estruturas secundárias das sequências de 50 ribonucleotídeos que contêm os TPS internos. A) Agrupamento em função das suas estruturas secundárias. B) Alinhamentos das sequências nos nós com 5 (verdes), 6 (azul), 7 (laranja), 8 (vermelho), 9, (amarelo), 10 (magenta), 11 (azul-claro) sequências, organizadas da esquerda para direita no sentido da transcrição 5'-3', com os respectivos valores de mínima energia livre (mfe) em kcal/mol e as estruturas consensos em formato “*Dot-bracket*”, onde os parênteses indicam bases pareadas e os pontos indicam as despereadas. Em linhas descontínuas o possível TPS com uma margem de 2nt de erro na posição.

Figura 4.41: Análises das estruturas secundárias das sequências de 50 ribonucleotídeos que contêm os TPS antisense. A) Agrupamento em função das suas estruturas secundárias. B) Alinhamentos das sequências nos nós com 5 (verdes), 6 (azul), 7 (laranja) sequências, organizadas da esquerda para direita no sentido da transcrição 5'-3', com os respectivos valores de mínima energia livre (mfe) em kcal/mol e as estruturas consensos em formato “*Dot-bracket*”, onde os parênteses indicam bases pareadas e os pontos indicam as despareadas. Em linhas descontínuas o possível TPS com uma margem de 2nt de erro na posição.

4.6 Expressão da proteína recombinante do gene VNG_RS05855 em E. coli

Para analisar a função in vitro da proteína codificada no gene VNG_RS05855 de *H. salinarum* NRC-1, geramos a proteína recombinante a partir da sua expressão heteróloga em *E. coli*, na busca por evidências experimentais diretas da atividade do seu produto proteico.

Figura 4.42: Gel de poliacrilamida 15% com as amostras do processo de purificação da proteína recombinante do gene VNG_RS05855 expressa em *E. coli* BL21 com o vetor pET-15b. Canaletas: 1) Marcador de peso molecular (Bio-rad); 2) Extrato total sem indução da expressão da proteína recombinante; 3) Extrato total com indução da expressão da proteína recombinante; 4) Fração *pellet* do processo de purificação; 5) Fração sobrenadante do processo de purificação; 6) Fração da proteína purificada por cromatografia líquida em coluna de níquel (Ni-NTA).

Após a expressão da proteína recombinante na linhagem BL21 de *E. coli* realizamos a avaliação das etapas do processo de purificação utilizando gel de poliacrilamida 15% (SDS-PAGE). Análises das diferentes frações do processo de purificação mostram que a proteína foi expressa de maneira eficiente como proteína solúvel e tem uma migração aproximada de 37 kDa, correspondendo com a massa molecular calculada para ela de 36.08 kDa (Figura 4.42). Esta fração foi quantificada utilizando o método colorimétrico Bradford e mostrou que a proteína está com uma concentração de 820 ng/ μ L.

4.6.1 Os conjuntos dos RNA das linhagens não são claramente diferentes

Antes do tratamento do RNA com a proteína heteróloga, analisamos a qualidade do conjunto dos RNA total e pequeno, previamente purificados das linhagens. Com esse objetivo, realizamos a eletroforeses de nano-capilares de 10 ng dos RNA usando do Bioanalyzer Agilent 2100 o chip RNA-Nano para o RNA total e o chip RNA-Small para o RNA total e pequeno.

Analisamos o RNA total purificado de duas réplicas biológicas das distintas linhagens. O perfil dos eletroferogramas das réplicas é apresentado na Figura 4.43. Da esquerda para direita os picos representam o seguinte: o primeiro, o marcador de peso molecular, em seguida e em forma de uma pequena meseta, observamos as regiões dos picos que definem os pequenos RNAs (incluindo os tRNAs), esta região é limitada pelo pico que representa o RNA ribossomal

Figura 4.43: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA das linhagens controle (Δ ura3-vermelho-) e nocaute (Δ VNG_RS05855-preto-). Apresenta-se o padrão de bandas do gel associado a cada corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). Eletroforeses do RNA total usando o chip RNA-Nano. A) Réplica 1 e B) Réplica 2

5S. Os picos maiores correspondem aos RNA ribossomais 16S e 23S, respectivamente, e um conjunto com os picos caracterizados anteriormente são fundamentais no eletroferograma do RNA total. Os picos restantes não são rotulados com especificidade (Technologies, 2005).

Distinguimos diferenças nos perfis dos eletroferogramas das linhagens controle e nocaute nos picos correspondentes aos tRNA e rRNA em uma das réplicas (Figura 4.43 B), porém o comportamento dessas diferenças não é reproduzível na outra réplica (Figura 4.43 A). Um aspecto que dificulta a comparação são as variações nas intensidades de fluorescência (FU) apesar de ter utilizado a mesma quantidade de RNA (10 ng). No entanto, apoiados na visualização da eletroforeses que acompanha ao eletroferograma percebemos que altura das bandas correspondentes ao rRNA e as intensidades não são visualmente diferentes entre as réplicas e cumprem os padrões de serem o rRNA 23S mais intenso que rRNA 16S (Technologies, 2006). A soma destes resultados nos permite sugerir que os perfis dos RNA total das linhagens não são facilmente distinguíveis.

Usando o chip RNA-Small específico para RNA pequeno, avaliamos a qualidade deste tipo de RNA previamente purificado em duas réplicas das linhagens controle e nocaute. Os perfis dos eletroferogramas representam da esquerda para direita: o marcador de peso molecular,

Figura 4.44: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA das linhagens controle (Δ ura3-vermelho-) e nocaute (Δ VNG_RS05855-preto-). Apresenta-se o padrão de bandas do gel associado a cada corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). Eletroforeses do RNA pequeno de duas réplicas biológicas de cada linhagens usando o chip RNA-Small. A) Réplica 1 e B) Réplica 2

as regiões de picos que correspondente aos tRNAs e pequenos-RNAs e por último o pico que caracteriza o rRNA 5S (Figura 4.44)(Technologies, 2016).

Entre as linhagens observamos a semelhança nos perfis dos tRNA entre as réplicas e as diferenças concentradas nas regiões que caracterizam o rRNA 5S, sendo reproduzíveis entre as réplicas e também verificável nos resultados no gel da eletroforeses (Figura 4.44). Este tipo de RNA está localizado como parte de uma unidade policistrônica que contém os genes que codificam para os RNA ribossomal 23S (rRNA 23S) e transportador Ala (tRNA-Ala). Em *E. coli* está caracterizada a liberação do rRNA5S da sua unidade policistrônica pela atividade da RNase E (Apiron e Miczak, 1993), porém em *H. salinarum* NRC-1, não se conhece a enzima que realiza este processo, apesar da proposta de uma regulação pós-transcricional (Clouet-d’Orval, 2017). O resultado poderia sugerir o apoio do produto proteico do gene VNG_RS05855 na maturação dos pré-rRNA 5S em *H. salinarum* NRC-1.

Com o chip RNA-Nano avaliamos a qualidade do RNA total das linhagens a partir de amostras de RNA total e com o chip RNA-Small os RNAs pequenos a partir de amostras de RNA pequeno. Porém, é possível avaliar a qualidade do RNA pequenos a partir de amostras

Figura 4.45: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA das linhagens controle (Δ ura3-vermelho-) e nocaute (Δ VNG_RS05855-preto-). Apresenta-se o padrão de bandas do gel associado a cada corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). A) Eletroforeses do RNA total de uma réplica biológica para cada linhagem usando o chip RNA-Small

de RNA total com o chip RNA-Small (Technologies, 2016). Ao realizar este experimento, com uma réplica biológica de cada linhagem, observamos que a principal característica do eletroferograma é maior intensidade dos picos na linhagem nocaute que na linhagem controle (Figura 4.45). O perfil de intensidades de cada linhagem é diferente ao que previamente tinha sido obtido ao estudar os pequenos RNA com o chip RNA-Small, especificamente na região que caracteriza o rRNA 5S (Figura 4.44). Uma possível explicação para este achado poderia ser a origem da amostra que estamos avaliando. Ao examinar o RNA total com o chip RNA-Small não avaliamos unicamente os RNA pequeno, ao serem amostras complexas de RNA além dos tRNA e pequenos-RNA podem existir outros RNA de tamanhos similares a eles originados ou do processamento ou da degradação. As diferenças entre as linhagens poderia indicar algum efeito da ausência do gene VNG_RS05855 no metabolismo do RNA. No entanto, as diferenças observadas entre as linhagens nessas regiões não são reprodutíveis avaliar o RNA total (Figura 4.43).

A avaliação da qualidade do conjunto dos RNA total e pequeno das linhagens mostrou resultados contraditórios. Perfis de fluorescência ou da intensidade da banda (eletroforeses) não reprodutíveis entre experimentos nos sugere que as diferenças entre os RNA das linhagens não

sejam consideráveis.

4.6.2 A proteína recombinante processa substratos de RNA in vitro

Para testar atividade da proteína heteróloga, ela foi incubada separadamente com o RNA total e o RNA pequeno das linhagens controle e nocaute no tampão recomendado (Späth et al., 2008). Após o tempo de incubação, a reação foi detida mediante extração com fenol/clorofórmio e os RNAs analisados por eletroforeses nano-capilar. Os produtos da reação da proteína com o RNA total foram analisados com o chip RNA-Nano e chip RNA-Small, os produtos da reação obtidos do RNA pequeno com chip RNA-Small.

Observa-se no eletroferograma do RNA total de cada linhagem após da sua incubação com a proteína recombinante a mudança de perfil de intensidade de fluorescência (FU) em relação às amostras não tratadas com a proteína, especificamente a perda das bandas correspondentes ao RNA ribossomal (Figura 4.46). Este resultado evidencia experimentalmente a atividade ribonucleolítica in vitro da proteína codificada no gene VNG_RS05855. O resultado também mostrou o enriquecimento das regiões correspondentes aos tRNAs ou pequenos-RNAs nos eletroferogramas. Este enriquecimento poderia ser a consequência de novos fragmentos de RNA originados a partir do processamento dos rRNA por parte do produto proteico do gene VNG_RS05855 que não alcança a processá-los na sua totalidade.

Ao avaliar os perfis de FU e o padrão de bandas da eletroforese após do tratamento do RNA pequeno com a proteína recombinante também observamos mudanças drásticas em relação a amostra não tratada (Figura 4.47). As maiores mudanças foram observadas na linhagem nocaute (Figura 4.47 B). O resultado sugere que o processamento seja preferencial a RNA destas amostras. Porém, nas amostras de RNA da linhagem nocaute também ocorre processamento. Interessante destacar o perfil do rRNA 5S após da incubação com a proteína recombinante, onde se consegue observar a diminuição da FU, sugerindo a capacidade da proteína para processar este transcrito.

Também estudamos o comportamento do RNA total após do tratamento com a proteína recombinante com o chip RNA-small. No perfil de FU das amostras da linhagem nocaute observamos variações na região do eletroferograma correspondente aos pequenos RNAs e tRNAs (Figura 4.48). Observa-se também como nas amostras tratadas com a proteína aparecem sinais

Figura 4.46: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA incubado com a proteína heteróloga do gene VNG_RS05855. Apresenta-se o padrão de bandas do gel associado com a corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). Eletroforeses usando o chip RNA-Nano do RNA total das linhagens controle (A) e nocaute (B) tratados com a proteína heteróloga (preto) e com seu tampão de reação (cinza)

Figura 4.47: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA incubado com a proteína heteróloga do gene VNG_RS05855. Apresenta-se o padrão de bandas do gel associado com a corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). Electroforese usando o chip RNA-Small do RNA pequeno das linhagens controle (A) e nocaute (B) tratados com a proteína heteróloga (preto) e com seu tampão de reação (cinza).

Figura 4.48: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares do RNA incubado com a proteína heteróloga do gene VNG_RS05855. Apresenta-se o padrão de bandas do gel associado com a corrida (nt: tamanho de nucleotídeos; L: bandas do ladder). Eletroforeses usando o chip RNA-Small do RNA total das linhagens controle (A) e nocaute (B) tratados com a proteína heteróloga (preto) e com o seu tampão de reação (cinza).

alterando os perfis clássicos do eletroferograma obtidos com o chip RNA-Small, sugerindo que sejam resultado do processamento realizado pela proteína sobre o RNA ribossomal, indicando que a atividade da proteína possa ser simultânea: RNA pequenos e ribossomais.

Avaliamos o efeito do tampão de reação da proteína heteróloga sobre a integridade do RNA total das linhagens controle e nocaute. Para isso, comparamos os eletroferogramas do RNA total das linhagens controle e nocaute, com e sem, o tampão de reação pelo mesmo tempo de incubação do RNA total com a proteína recombinante.

Observamos, independentemente da linhagem, a inversão da intensidade de FU dos picos do rRNA 16S e 23S (Figura 4.49). A maior intensidade do pico do rRNA 16S em relação ao 23S é indicativo da degradação parcial do RNA total, sugerindo que o tampão pode afetar a integridade dos RNA. No entanto, ao observar o efeito da proteína recombinante sobre esse RNA (Figura 4.46) sugerimos que a influência do tampão sobre a integridade dos RNA não seja suficiente como para justificar o nível de processamento observado quando o RNA total é tratado com a proteína recombinante.

Figura 4.49: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares e o padrão de bandas do gel associado (nt: tamanho de nucleotídeos; L: bandas do ladder) do RNA total das linhagens controle (A) e nocaute (B) sem tratamento (vermelho e preto, respectivamente) e tratados com o *tampão* de reação da proteína heteróloga (cinzas) analisadas com o chip RNA-Nano

4.6.3 Perfil do processamento da proteína recombinante é similar ao das RNases R e A

Por último examinamos os perfis de intensidade de FU e o padrão de bandas da eletroforese obtidos do tratamento do RNA total e pequeno da linhagem nocaute com RNases comerciais com atividade conhecida (RNase A, RNase R, RNase V1, RNase S1). O objetivo deste experimento foi comparar os perfis do processamento destas enzimas com o perfil obtido para proteína recombinante rVNG_RS05855.

Nós testamos a RNase A, endonuclease que digere RNA de fita simples e cliva especificamente após nucleotídeos de pirimidina (ThermoScientific, 2017); RNase R, digere todos os RNAs lineares, mas não digere as estruturas de RNA circular ou RNA de cadeia dupla com saliências menores que sete nucleotídeos na extremidade 3' (Epicentre, 2012); RNase S1, capaz de degradar RNA de fita simples na cadeia linear ou originada no interior do laço (Sigma, 2000) e RNase V1, endonuclease que cliva dupla fita de RNA (Grade, 2012).

O RNA total (10 ng) tratado com a RNase A, a RNase R, a RNase S e a RNase V1 mostrou a degradação completa do rRNA e picos de intensidade de FU desprezíveis nas regiões

Figura 4.50: Unidades arbitrárias de fluorescência em função do tempo da corrida eletroforética em nano-capilares e o padrão de bandas do gel associado (nt: tamanho de nucleotídeos; L: bandas do ladder) A) Do RNA total incubado com enzimas ribonucleolíticas de atividade conhecida (RNases A, R, S, V1) analisado com chip RNA-Nano; B) Do RNA pequeno incubado com enzimas ribonucleolíticas de atividade conhecida (RNases A, R, S, V1) analisado com chip RNA-Small.

que caracterizam os tRNA e pequenos-RNA (Figura 4.50 A). Com este resultado sugerimos que degradação do RNA total por parte da proteína recombinante rVNG_RS05855 é menos eficiente que a realizada pelas RNases avaliadas.

Ao avaliar o efeito das RNases caracterizadas na degradação do RNA pequeno observamos que as RNases S1 e V1 degradam totalmente o RNA pequeno, enquanto as RNases A e R o degradam parcialmente (Figura 4.50 B). Os perfis da atividade da proteína recombinante rVNG_RS0588 sobre os RNA pequenos parecem ser uma mistura dos perfis de degradação das RNase A e R (Figura 4.47), sugerindo que a proteína recombinante possa-se comportar como elas, sendo endo e exonuclease.

5 Discussão

5.1 Características da sequência VNG_RS05855 sugerem que seu produto proteico seja uma RNase

O primeiro passo na caracterização do gene VNG_RS05855 de *Halobacterium salinarum* NRC-1 foram as análises da sua sequência proteica (Figura 4.1) e o estudo da sua expressão relativa em diferentes backgrounds genéticos e condições ambientais (Figura 4.2 e 4.3).

Na sua sequência proteica identificamos os cinco motivos conservados que caracterizam a família proteica das metalo- β -lactamase e a ausência do exosite, domínio que permite o reconhecimento do pré-tRNAs. No entanto, mostramos que estas características são compartilhadas com a região carboxi terminal da RNase Z^L de *S. cerevisiae* que mantêm as propriedades ribonucleolíticas (Ma et al., 2017; Saoura et al., 2017). O perfil de expressão do RNA de VNG_RS05855 não apresentou grandes variações, sugerindo um papel constitutivo. Observamos a correlação da sua expressão com a de outras RNases, o que nos indicou uma possível ação redundante e coordenada da sua atividade, possivelmente promovendo modificações no conjunto de transcritos que garantam o posterior processamento por outras RNases.

Uma proteína de *Haloferax volcanii* que compartilha 66% de identidade de sequência com a codificada por VNG_RS05855 foi estudada por outros pesquisadores (Fischer et al., 2012). A proteína codificada por HVO_RS18055 (anotação antiga HVO_2763) mostrou ser uma metalo- β -lactamase (MBL). Entre os membros das MBL, além das ribonucleases, estão as glioxalases II, arilsulfatases e fosfodiesterase. Testes para comprovar as três últimas atividades da proteína codificada no gene HVO_RS18055 (NZ/Hvo) foram realizadas, porém nenhuma dessas atividades foi observada (Fischer et al., 2012). Com base na similaridade de sequências entre esta proteína de *H. volcanii* e a por nós estudada em *H. salinarum* NRC-1, sugerimos como pouco provável que o produto proteico do gene VNG_RS05855 tenha as funções das glioxalases II, arilsulfatases e fosfodiesterase.

A função do produto proteico NZ/Hvo de *H. volcanii* também foi avaliada em conjunto com a atividade da RNase Z do mesmo organismo mediante a incubação in vitro de ambas proteínas com um grupo de pré-tRNA e pré-rRNA 5S. Os resultados mostraram que a adição de

NZ não teve efeito visível na atividade da RNase Z (Fischer et al., 2012). É possível que VNG_RS05855 tenha atividade semelhante a NZ e não descartamos a possibilidade de ações coordenadas com outras RNases. Por exemplo, no desligamento da expressão de membros de operons com autorregulação positiva, fenômeno já caracterizado em *H. salinarum* NRC-1, é proposto que o processamento do ativador transcricional seja realizado por um conjunto de RNases (Wurtmann et al., 2014). De uma maneira similar, propõe-se que frente a mudanças na disponibilidade de oxigênio, as alterações na expressão de genes relacionados com a fermentação de arginina, respiração anaeróbica (TMAO e DMSO, como aceptores finais dos elétrons), produção das vesículas de gás e de citocromos oxidases podem estar relacionados a possível ação de RNases, apesar dos mecanismos permanecerem desconhecidos (DasSarma et al., 2012). Sendo assim, e com a aceitação de que a ação coordenada e redundante é característica das atividades das RNases (Redko, Li de Lasierra-Gallay e Condon, 2007; Deutscher, 2015; Clouet-d'Orval, 2017), podemos sugerir que poderiam existir vários processos em *H. salinarum* NRC-1 onde a participação do produto proteico do VNG_RS05855 esteja relacionada com outras RNases e deveriam ser explorados em pesquisas futuras.

A expressão das RNases é estritamente regulada durante todo o crescimento celular e influenciada pelo contexto ambiental (Wurtmann et al., 2014; Deutscher, 2015). Em *E. coli* estudos da expressão de RNases mostraram ter um perfil constitutivo durante o crescimento e que se modifica em função das mudanças ambientais e genéticas (Schilling et al., 2004). Estas características são reprodutíveis, em nossos resultados, ao avaliar a expressão relativa do conjunto de RNases de *H. salinarum* NRC-1 e são características compartilhadas com o transcrito do gene VNG_RS05855.

5.2 A ausência do gene VNG_RS05855 não afeta o crescimento da linhagem nocaute e influencia de forma sutil a resposta perante estresses

A linhagem nocaute do gene VNG_RS05855 mostrou ser viável e sem diferença no crescimento em relação a linhagem controle em condições de referência do crescimento. Observações semelhantes foram obtidas em *H. volcanii* e *E. coli*, onde linhagens nocaute de genes que compartilham 66% e 31% de identidade de sequência proteica como a sequência em estudo,

respectivamente, foram estudados e tampouco mostraram fenótipos (Schilling et al., 2004; Fischer et al., 2012).

No entanto, pequenas diferenças fenotípicas são observadas entre as linhagens nocaute e a linhagem controle perante as condições estressantes: metais de transição, metalóide, um agente oxidante e mudanças na concentração de sais do meio de cultura.

O produto proteico do gene VNG_RS05855 poderia ter um efeito sobre transcritos de transportadores de metais de transição, como por exemplo: Cu^{2+} (NosFY), Fe^{2+} (permease de ferro HemUV), Zn^{2+} (ZurMA) e Co^{2+} (CbiNOQ), etc (Ng et al., 2000). Contudo, as respostas a metais de transição em *H. salinarum* NRC-1 também envolve uma rede de treze distintos reguladores transcricionais. O estudo de dois deles permitiu confirmar que são capazes de atuar indistintamente sobre a expressão de genes que participam no transporte de diferentes metais (Kaur et al., 2006). Neste contexto, o efeito da ausência do gene VNG_RS05855 poderia relacionar-se não só ao transcritos de proteínas efetoras do transporte, também a proteínas reguladoras da expressão. Por outro lado, observamos que com os aumentos nas concentrações dos metais de transição a tendência no comportamento das diferenças de OD_{600nm} das linhagens é aproximar-se a zero (Figura 4.10 B, C, D). Isto sugere que aumentos das concentrações dos metais de transição levam a ativação de mecanismos adaptativos comuns entre as linhagens, sem a eventual participação do produto proteico do gene VNG_RS05855.

Diferente é a resposta da linhagem nocaute frente ao metalóide arsenito de sódio (NaAsO_2), sendo os valores das diferenças de OD_{600nm} entre as linhagens sempre negativas, indicando a susceptibilidade da linhagem nocaute frente ao arsenito (Figura 4.10 H). Em *H. salinarum* NRC-1 os genes de resistência a arsênico, arsADRC, estão maioritariamente localizados no plasmídeo pNRC100 e são membros de um operon divergente (Ng et al., 2000). Em *E. coli* foram descritos processamentos pós-transcricionais no operon de resistência a arsenito para assegurar a correta resposta frente ao estresse (Owolabi e Rosen, 1990); sugerimos que diante da ausência de uma possível RNase em *H. salinarum* NRC-1 processamentos similares poderiam ser omissos e resultar no aumento da susceptibilidade da linhagem nocaute.

Diante do agente oxidante H_2O_2 a resposta da linhagem nocaute parece ser contraditória. Na menor concentração do agente oxidante a ausência do gene VNG_RS05855 constitui

uma desvantagem na linhagem nocaute e nas maiores concentrações representa uma vantagem (Figura 4.10 I). Em *E. coli* está descrito que a expressão de genes chaves nas respostas ao estresse oxidativo evidenciam a redução significativa da sua expressão na ausência de RNases (Touati e Dubrac, 2015). Este seria um evento que também poderia ocorrer na linhagem nocaute do gene VNG_RS05855, evidenciando a maior sensibilidade. Por outro lado, estudos relatam que concentrações maiores de 7 mM de H_2O_2 retardam o crescimento de *H. salinarum* NRC-1 de maneira permanente (Kaur et al., 2010). O fenótipo no qual a célula atrasa seu crescimento e não consegue restabelecê-lo é conhecido com o estado de parada do crescimento (“growth arrest”). Neste estado a célula ajusta a maquinaria de transcrição mediante longas pausas e lentas taxas de alongação (Bergkessel, Basta e Newman, 2016). Nestas condições, a ausência de uma possível ribonuclease na linhagem nocaute poderia aumentar a estabilidade do baixo número de RNA produzidos, sugerindo a vantagem da linhagem nocaute frente a maior concentração do H_2O_2 .

Também se distinguem ligeiras diferenças fenotípicas entre as linhagens ao diminuir a concentração de sais do meio (Figura 4.10 J). Adaptação à salinidade por parte de arqueias halofílicas requer mudanças regulatórias e metabólicas para a manutenção do balanço osmótico (Beer, Wurtmann e Baliga, 2014). *H. salinarum* NRC-1 garante sua homeostase mediante o acúmulo de K^+ ou mediante a síntese e/ou acúmulo de solutos orgânicos osmoticamente ativos capazes de excluir o sal do citoplasma (ex: açúcares, derivados de aminoácidos) (Oren, 2008). Neste organismo já têm sido descritos modificações pós-transcricionais nos transcritos dos operon de captação de K^+ (Wurtmann et al., 2014). Por outro lado, a captação de metabólitos osmoticamente ativos em *H. salinarum* NRC-1 é realizada por membros da superfamília de transportadores ABC (Ng et al., 2000). Em *B. subtilis* e *S. enterica* foi mostrado que membros destas famílias de transportadores sofrem modificações pós-transcricionais (Allenby et al., 2004; Miyakoshi, Chao e Vogel, 2015). Com estes antecedentes, nossos resultados podem indicar que a susceptibilidade da linhagem nocaute frente a mudanças dos sais dos meios pode relacionar-se a ausência de possíveis modificações pós-transcricionais nos transportadores de produtos osmoticamente ativos.

As diferenças fenotípicas entre as linhagens associadas ao defeito na regulação pós-

transcricional como consequência das diferenças entre os níveis de mRNA e proteínas foi previamente proposto por vários autores (Trey Ideker, Vesteyinn Thorsson, 2001; Baliga et al., 2002). Com esta base, explicamos as ligeiras diferenças fenotípicas encontradas entre as linhagens controle e nocaute. Contudo, existem discrepâncias com este enfoque e se propõe que as diferenças entre os níveis de mRNA e proteínas poderia ser pela produção diferencial de proteína a partir de um único mRNA e não da regulação pós-transcricional (Kaur et al., 2006; Kaur et al., 2010). No entanto, mediante modelagem matemática se demonstrou que o nível de proteína se regula mais rapidamente a nível pós-transcricional, seguido da pós-traducional e via de regulação transcricional como a abordagem mais lenta (Levine et al., 2007; Shimoni et al., 2007; Mehta, Goyal e Wingreen, 2008; Wurtmann et al., 2014). Assim, podemos presumir que a produção de mais proteína a partir de um mRNA transitório é possível se mecanismos de regulação operam em conjunto para seu favorecimento; dentre eles, o mecanismo pós-transcricional é o mais valorizado para assegurar graus diferentes entre proteínas e mRNA na maior brevidade. Uma explicação para o uso não uniforme da regulação pós-transcricional, pós-traducional e transcricional é que a interação seletiva desses mecanismos constitui uma estratégia evolutiva para gerar respostas dinâmicas adequadas perante perturbações ambientais (Wurtmann et al., 2014).

5.3 Na linhagem nocaute a produção de bacteriorodopsina não é afetada, porém diante de mudanças na disponibilidade de nutrientes, as transições metabólicas são perturbadas

Os incrementos na OD_{600nm} nas fases finais do crescimento podem indicar o possível efeito do gene VNG_RS05855 com fatores que o promovem: a produção de bacteriorodopsina e/ou a liberação de vesículas de gás, e em menor medida ao crescimento celular. No conjunto, esses fatores incrementam a dispersão da luz em *H. salinarum* NRC-1 (Facciotti et al., 2010). Contudo, os estudos de cada fator permitiu sugerir que entre as linhagens não existem diferenças entre as densidades celulares nem na produção de bacteriorodopsina, nem na produção interna de vesículas de gás. Essa diferença foi associada com a alta dispersão dos dados de absorbância das linhagens de *H. salinarum* NRC-1, seguindo a sugestão feita por (Tonner et al., 2017).

A bacteriorodopsina pode promover mudanças nas medições de absorbância e, junto

com a respiração celular, é um dos mecanismos de obtenção de energia química que promove mudanças no pH celular (Oesterhelt e Krippahl, 1973; Oesterhelt e Stoeckenius, 1973). As mudanças no pH ocasionadas pela bacteriorodopsina são influenciadas pela presença de luz e a disponibilidade de nutrientes, de maneira similar como ocorre na respiração celular, no entanto, nesta última, a luz não tem efeito (Baliga et al., 2002).

Ao promover a restrição transitória de nutrientes com diluições de solução basal em diferentes fases do crescimento, observamos que as variações de pH entre as linhagens aparecem na fase de crescimento exponencial e estacionária, só na presença da luz (Figura 4.13 A). Na fase de crescimento exponencial, os principais transcritos encontrados no crescimento de *H. salinarum* NRC-1 estão relacionados fundamentalmente a maquinaria de transcrição, tradução e fosforilação oxidativa. Nesta fase, nas que as células se encontram às 41 horas, a linhagem nocaute deixa o meio extracelular mais básico que a linhagem controle. A alcalinização do meio extracelular pode ser promovida por proteínas que facilitam a internalização de prótons como transportadores antiporte ou ATP sintetase. O transportador antiporte descrito em *H. salinarum* NRC-1 que participa no efluxo de Na⁺ e é o NhaC (Ng et al., 2000) e não achamos na literatura consultada referências a possíveis mecanismos pós-transcricionais que regulam a sua expressão em *H. salinarum* NRC-1. A ATP sintase, por sua vez, torna o meio extracelular mais básico mediante a internalização de prótons, o que conduz a formação de ATP (Grüber et al., 2014). Mecanismos de regulação pós-transcricional têm sido descritos para os transcritos destas enzimas em mitocôndrias (Willers e Cuezva, 2011; Long, Yang e Yang, 2015; Izquierdo et al., 1995). Na limitação de nutrientes, a alcalinização do meio extracelular na fase de crescimento exponencial poderia sugerir que a linhagem nocaute esteja produzindo mais ATP. A diferença não representa um drástico efeito no crescimento das linhagens, contudo as mudanças nessa faixa de pH não afeta o crescimento de *H. salinarum* NRC-1 (Moran-Reyna e Coker, 2014).

Às 95 horas do crescimento, na fase de crescimento estacionária (Figura 4.13 A), o meio extracelular da linhagem nocaute é mais ácido que na linhagem controle. Nesta fase, o esgotamento de nutrientes e a queda na disponibilidade de oxigênio explica a expressão de uma maior quantidade de genes envolvidos nos processos alternativos de obtenção de energia (Lange et al., 2007). O papel da bacteriorodopsina nestas condições é garantir a manutenção do gra-

diente eletroquímico mediante a exportação de prótons ao meio extracelular que são utilizados na produção de ATP, visto que o consumo dos nutrientes esgotou as moléculas com poder redutor na célula. A ausência de uma possível ribonuclease poderia causar a acidificação do meio extracelular pelo aumento do tempo de meia vida do RNA da bacterioopsina.

Em bactérias se tem comprovado que a expressão de algumas RNases podem ocorrer em fases específicas do crescimento e ser influenciadas pelo contexto ambiental (Deutscher, 2006; Deutscher, 2015). Mecanismos similares não foram ainda mostrados nas RNases de Archaea, precisamente pelo pouco conhecimento das enzimas ribonucleolíticas nos organismos deste domínio, contudo, é possível que existam e apoiem a explicação dos fenótipos das diferenças de pH encontrados entre as linhagens controle e nocaute.

Ao disponibilizar todos os nutrientes em diferentes fases do crescimento nas linhagens mediante diluições com meio CM, observamos que não existem diferenças significativas nas medições do pH na luz (Figura 4.14 A). Supomos que a presença de vários mecanismos de obtenção de energia nestas condições equilibrem as demandas energéticas das células e mantenha o pH inalterável entre as linhagens. No entanto, no escuro, há diferenças no pH entre as linhagens às 17, 65 e 95 horas do crescimento (Figura 4.14 B). No escuro não ocorre fototrofia, e o processo que pode alterar o pH é a respiração, indicando que VNG_RS05855 esteja relacionado com os transcritos associados ao transporte de elétrons. Durante a respiração, esses genes se encarregam de manter o gradiente protônico mediante a movimentação de prótons ao meio extracelular o que, além de provocar a acidificação, favorece a formação de ATP (WAGNER, HARTMANN e OESTERHELT, 1978; Talaue et al., 2016). Contudo, as mudanças no pH se mantém na faixa que não provoca variações no crescimento de *H. salinarum* NRC-1 (Moran-Reyna e Coker, 2014).

A importância da regulação pós-transcricional nos mecanismos de obtenção de energia em *H. salinarum* NRC-1 já foi sugerida e foram caracterizadas como eventos claramente complexos e que permanecem pouco compreendidos (Baliga et al., 2002). A proposta que vários processos de obtenção de energia operem simultaneamente como uma ferramenta adaptativa para encarar a disponibilidade irregular de nutrientes (Gonzalez et al., 2009) reforça a interconexão e importância dos mecanismos de regulação, nele os resultados sugerem que produto

proteico do gene VNG_RS05855 desempenhe algum papel.

5.4 A ausência gene VNG_RS05855 não perturba a produção de carotenóides nem vesículas de gás

É conhecido que a montagem do principal componente da fototrofia, a bacteriorodopsina, está co-regulada com as vias de síntese de carotenóides e isoprenóides (Baliga et al., 2002). O estudo da produção diferencial desses elementos entre as linhagens foi realizado mediante a espectroscopia Raman.

Os principais resultados mostraram que os espectros Raman de ambas linhagens são dominados pelas características dos picos dos carotenóides as 1002 cm^{-1} , 1152 cm^{-1} e 1507 cm^{-1} que pode ser atribuído a bacterioruberina, carotenóide de 50 átomos de carbono (C_{50}) com 13 duplas ligações conjugadas (Marshall et al., 2006; Leuko et al., 2007). Na vizinhança desses picos também existem outros como 1008 cm^{-1} , 1157 cm^{-1} , 1515 cm^{-1} , 1520 cm^{-1} que correspondem a uma mistura de carotenóides que apresentam variações no número de suas duplas ligações conjugadas na cadeia principal e poderiam ser desde carotenóides de C_{40} até β -carotenos ou licopenos (Fendrihan, Musso e Stan-Lotter, 2009). No entanto, nem na bacterioruberina nem na mistura de outros carotenóides distinguimos diferenças significativas na composição entre as linhagens nos distintos comprimentos de onda em que se excitaram as amostras extraídas das diferentes fases do crescimento das linhagens (Figura 4.15 e 4.16).

Observamos um comportamento na fração espectral entre 1250 cm^{-1} - 1450 cm^{-1} não caracterizado previamente em *H. salinarum* NRC-1. Nesta região, que se associa aos ácidos graxos de membranas de eucariotos e bactérias, geralmente nas arqueias têm intensidades Raman fracas (Pande, Lanyi e Callender, 1989; Leuko et al., 2007; Fendrihan, Musso e Stan-Lotter, 2009). O resultado poderia indicar que a composição das membranas esteja sendo afetada, porém não são resultados reprodutíveis só na linhagem nocaute, aparecendo intensidades Raman nesta fração espectral também na linhagem controle. A interpretação deste resultado deve ser reservada após confirmar o resultado com a utilização de um maior número réplicas do experimento.

A busca de diferenças entre as linhagens associadas as intensidades Raman de diferentes compostos fotossensíveis como a bacteriorodopsina ou halorodopsina não evidenciou

diferenças nos espectros Raman (Figura 4.17). No entanto, os incrementos nas absorbâncias nas etapas finais do crescimento também poderiam estar relacionados com a liberação ao meio extracelular de vesículas de gás (Facciotti et al., 2010).

A inspeção visual de vesículas de gás no interior das células entre as linhagens não mostrou claras diferenças (Figura 4.18). Juntamente com o fato de não haver diferença na OD_{600nm}, estes fatos sugerem um grau semelhante de estabilidade celular e consequentemente, número semelhante de vesículas de gás liberados no meio extracelular. Assim, provavelmente as vesículas não sejam a causa dos incrementos observados nas diferenças OD_{600nm} nas etapas finais do crescimento.

5.5 VNG_RS05855 não desempenha um papel-chave na sobrevivência de *H. salinarum* NRC-1 frente a Cu²⁺, AsO₂⁻ e PQT

Os ensaios de sobrevivências permitiram estudar as diferenças entre a viabilidade celular das linhagens controle e nocaute após das células serem submetidas a concentrações semi-letais de agentes estressantes. Este experimento também permitiu confirmar, mediante a contagem de unidades formadoras de colônias (UFC), que a viabilidade das células nas condições referência do crescimento não são diferentes entre as linhagens controle e nocaute.

Os ensaios de sobrevivência foram realizados com concentrações semi-letais dos agentes estressantes nos que se detectou diferenças na OD_{600nm} entre as linhagens: Cu²⁺ e AsO₂⁻. Além disso, avaliou-se a sobrevivência frente ao PQT, agente oxidante que causou diferenças nas OD_{600nm}, tendo o efeito mais drástico nas linhagens, como os menores valores de OD_{600nm} durante os crescimento.

A partir das medições OD_{600nm} no decorrer do crescimento entre as linhagens frente a Cu²⁺ e AsO₂⁻, observamos que com os aumentos nas concentrações destes agentes estressantes aumenta a inibição do crescimento. No entanto, ao com concentrações semi-letais tanto de Cu²⁺ como de AsO₂⁻ e promovendo o condicionamento prévio das células não é possível distinguir diferenças na viabilidade entre as linhagens (Figura 4.19 e 4.20). A enorme quantidade de dois agentes estressantes necessários para alcançar efeitos semi-letais é um testemunho da notável capacidade das linhagens de eliminar a concentração intracelular destes componentes e

mantê-las dentro de limites fisiológicos. Porém, o resultado nos sugere que produto proteico do gene VNG_RS05855 não participe nas respostas adaptativas das linhagens previamente acondicionadas aos estresses. Frente ao PQT pode-se comprovar este mesmo efeito porque apesar do drástico efeito sob o crescimento, diante o condicionamento prévio as células são capazes de enfrentar o dano por ele induzido sem diferenças na viabilidade (Figura 4.21), o que nos sugere que não há participação do gene VNG_RS05855.

5.6 A regulação por RNA antisense de genes relacionados ao operon das vesículas de gás pode ser afetada na ausência do gene VNG_RS05855

Para apoiar a caracterização fenotípica da linhagem nocaute do gene VNG_RS05855 comparamos o seu transcrito com o da linhagem controle. Com base na conhecida função repressora das RNases, prevemos que a deleção do gene VNG_RS05855 resultaria, predominantemente, no aumento da expressão de um maior número de genes, conforme ao sugerido por (Wurtmann et al., 2014) ao estudar a linhagem nocaute da RNase-VNG2099 de *H. salinarum* NRC-1. Em nossos resultados observamos padrão semelhante: maior o número de genes com expressão aumentada do que os que têm expressão diminuída. Na linhagem nocaute, 80 genes apresentam aumentos na sua expressão em relação à linhagem controle, sendo só 37 os que têm a expressão diminuída.

Entre os genes com expressão aumentada estão os relacionados a organização das vesículas de gás. No entanto, não observamos diferenças significativas entre as linhagens ao avaliar a produção interna destas estruturas supramoleculares.

Ao observar a cobertura de reads na linhagem nocaute do maior componente estrutural das vesículas de gás, o gene *gvpA*, distinguimos que além de apresentar níveis mais altos do que a linhagem controle, há a expressão de transcritos na fita oposta na extremidade 5'. Nesta região também se observa a maior cobertura de reads na linhagem nocaute que na linhagem controle (Figura 4.22). Os transcritos na fita oposta a um determinado gene são RNAs conhecidos como RNA não codificantes antisense, asRNA. Para o transcrito do gene *gvpA*, a existência do asRNA foi previamente descrito por nosso grupo em *H. salinarum* NRC-1 (Almeida et al., 2019). Pela localização próxima a extremidade 5', podemos sugerir que a formação do duplex mRNA-

asRNA tenha um papel na regulação da expressão por interrupção dos eventos de tradução. Esta atividade dos asRNA têm sido bem caracterizada (Sharma et al., 2007; Babski et al., 2014; Cech e Steitz, 2014; Xu, Zhang e Zhang, 2018) e em nossos resultados constitui uma explicação plausível para entender como os aumentos na expressão do maior componente estrutural das vesículas de gás na linhagem nocaute não implica incrementos internos na produção destas estruturas supramoleculares.

A origem do asRNA do transcrito *gvpA* se associa com possíveis eventos pós-transcricionais (Almeida et al., 2019). No entanto, nessa região não conseguimos distinguir possíveis sítios de processamento pós-transcricional (TPS) produto da atividade do VNG_RS05855. É possível sugerir a maior estabilidade do transcrito na linhagem nocaute de seu gene cognato mediante os incrementos na sua cobertura de reads. Conseguimos identificar um TPS no gene *gvpA*, sugerindo a maior estabilidade de seu mRNA na linhagem nocaute (Figura 4.35).

A degradação de mRNAs mediada por RNases em complexos ncRNAs-mRNA é uma característica da regulação da expressão em bactérias (Evguenieva-Hackenberg e Klug, 2011). A questão de saber se existe uma intersecção entre as RNases e os ncRNAs continua sendo um mistério em Archaea (Gelsinger e Diruggiero, 2018). Com nossos resultados, é possível supor que os asRNAs, além de deter a tradução, tenham alguma relação com a proteína codificada pelo gene VNG_RS05855.

A interrupção dos eventos de tradução e o aumento na estabilidade do transcrito do gene *gvpA* contribuem na interpretação dos aumentos da sua expressão, ainda quando o repressor da transcrição deste gene (*gvpD*) também tenha a expressão aumentada na linhagem nocaute. Porém, o resultado nos sugere que a repressão do operon por parte da proteína GvpD não é tão eficiente. Achados similares foram propostos por (Yao e Facciotti, 2011) que mostraram que existem condições nas quais a ausência de vesículas de gás ocorre independente da repressão pela proteína GvpD. Foi sugerido um mecanismo pós-transcricional de regulação similar com os elementos de resposta a ferro (IRE), porém com outras moléculas mediadoras, sem detalhar bioquimicamente o processo (Yao e Facciotti, 2011).

Outro membro do operon de vesículas de gás com expressão diferencial entre as linhagens é o gene *gvpO* (Figura 4.35). Neste gene observamos TPS localizados na fita oposta

a sua região codificadora, num asRNA também caracterizado em *H. salinarum* NRC-1 onde se identificou sinais de início da transcrição (Almeida et al., 2019). Assim como os asRNA podem bloquear a tradução, sua união a regiões codificadoras internas pode favorecer tanto estabilização do mRNA mediante a obstrução dos sítios alvos de endoribonucleases quanto favorecer a sua degradação, pela formação de fita dupla (asRNA-mRNA) que podem ser sinais para ação de endonucleases (Gelsinger e Diruggiero, 2018; Almeida et al., 2019). No entanto, o aumento na expressão do *gvpO* na linhagem nocaute sugere o favorecimento da estabilização do transcrito uma vez que o asRNA não apresentou TPS.

O papel da regulação antisense na expressão dos membros do operon de vesículas de gás já tinha sido sugerido (Krüger e Pfeifer, 1996; Almeida et al., 2019), porém não se conheciam os elementos ribonucleolíticos que poderiam estar relacionados. Com nossos resultados sugerimos a participação do VNG_RS05855 com essa função, como parte da desconhecida rede de regulação da expressão em *H. salinarum* NRC-1.

5.7 VNG_RS05855 pode estar envolvido na regulação da expressão de genes relacionados à família de transportadores ABC

Em *Haloferax volcanii* a linhagem nocaute do gene HVO_RS18055 (NZ/Hvo) mostrou que as principais mudanças na expressão gênica em relação a linhagem controle investigado por ensaios de microarray ocorreram na expressão do gene da sintase de triptofano e os relacionados com transportadores ABC (Fischer et al., 2012). Em nossos resultados, os membros da família de transportadores ABC têm a expressão alterada entre as linhagens nocaute e controle. O resultado poderia ser uma evidência da possibilidade de que os produtos proteicos destes genes tenham funções conservadas entre organismos.

Existem vários exemplos da regulação pós-transcricional mediada pela atividade de RNases em transportadores da família ABC. Um exemplo é o operon *pst* de *B. subtilis*, no qual a ação de exonucleases 3'-5' promove a formação de transcritos secundários. A regulação diferencial dos genes do operon ocorre pela interrupção desse processamento por estruturas "stem-loop" localizadas downstream ao primeiro gene do operon (Allenby et al., 2004). Ao observar o perfil de cobertura de reads para os membros do operon *dpp* (da família de transpor-

tadores ABC) com expressão aumentada na linhagem nocaute, conseguimos reconhecer sinais semelhantes aos descritos para *B. subtilis* (Figura 4.24 A). Os incrementos na cobertura de reads da extremidade 3' do operon na linhagem nocaute sugere que essa região seja alvo da possível atividade exoribonuclease 3'-5' do produto proteico do gene VNG_RS05855. A semelhança na cobertura de reads entre as linhagens no primeiro gene da unidade policistrônica poderia sugerir que a atividade ribonuclease não contínua nesta região, nas que se detectam TPS (Figura 4.34). Nesse sentido, a função da proteína poderia gerar transcritos alternativos mediante o controle dos níveis de cada membro do operon.

O mecanismo de regulação pós-transcricional descrito por (Allenby et al., 2004) em *B. subtilis* foi relatado para o operon da família de transportadores ABC relacionado ao transporte de fosfatos: *pst*. Em nossos resultados o operon *pst* de *H. salinarum* tem a cobertura de reads diferente da observada para o operon *dpp*. A expressão dos membros do operon *pst* é diminuída na linhagem nocaute em relação à linhagem controle (Figura 4.24 B). Ainda que a ação do produto proteico do gene VNG_RS05855 possa ser conservada em alguns operons, o resultado evidencia que a sua função varia de operon a operon. A diminuição da expressão de genes na linhagem nocaute indica que o metabolismo do RNA é muito mais complexo e o impacto da sua função na regulação pós-transcricional mais heterogêneo (Marujo et al., 2000; Qi et al., 2017).

Na arqueia metanogênica *Methanobrevibacter smithii* R15 mostrou-se que a remoção de determinadas estruturas “stem-loop” em regiões 5'-UTR dos transcritos não só promovem maior eficiência nos eventos de tradução como também incrementos na estabilidade do mRNA (Qi et al., 2017). Em bactérias, a ação de determinadas exoribonucleases promovem incrementos na estabilidade de transcritos (Marujo et al., 2000; Mohanty e Kushner, 2003; Pobre e Arraiano, 2015). Sendo assim, podemos associar a expressão diminuída de genes na linhagem nocaute a falta de processamento pós-transcricional. Entre os genes que supostamente precisam de processamento, e por essa razão estão com expressão diminuída na linhagem nocaute, além dos genes do operon *pst*, também encontramos os do operon de resposta a arsênico. Precisamente a diminuição na expressão destes genes na linhagem nocaute (Figura 4.23) pode ser a causa da sua sensibilidade na frente do arsênico, como foi evidenciado no ensaios fenotípicos realizados (Figura 4.7).

No entanto, não descartamos que a diminuição na expressão de genes na linhagem nocaute esteja relacionada ao estabelecimento de relações indiretas do VNG_RS05855 com reguladores transcricionais. Em nossos dados reguladores transcricionais também apresentam expressão aumentada na linhagem nocaute. Um exemplo, os fatores transcricionais MarR (Figura 4.25), que tem tanto função ativadora como repressora (Wilkinson). A função ativadora é exemplificada com os aumentos na expressão do gene VNG_RS01160 que está sob a sua regulação. No entanto, a ação repressora não é conhecida, basicamente por serem estes reguladores poucos compreendidos nas arqueias. Outros genes poderiam ter a sua expressão diminuída pela ação indireta dos reguladores transcricionais afetados pela ausência de VNG_RS05855.

Um resultado que apoia a relação do VNG_RS05855 com o processamento nos distintos operons dpp, pst e marR é a existência de TPS associados a ação ribonucleolítica (Figura 4.34 e 4.36, respectivamente). No entanto, a localização dos TPS varia nos genes dos operons com expressão diferencial, confirmando mais uma vez a ampla diversidade e complexidade de mecanismos de regulação pós-transcricionais que podem ser guiados pela atividade do produto proteico do gene VNG_RS05855.

5.8 A localização do produto proteico do gene VNG_RS05855 pode ser restrita a determinados espaços celulares

Um grande número de genes com expressão diferencial entre as linhagens, estimados pelo DESeq2, têm em comum a característica de serem seus produtos proteicos associados às membranas celulares segundo as suas anotações no RefSeq (Pruitt, Tatusova e Maglott, 2007)(Tabela 4.4, 4.5). Por exemplo, os transportadores Dpp, Pst, MATE, TrkA, EamA, Ars, membros da família de citocromos e fotossensores são genes com expressão diferencial entre as linhagens nocaute e controle associados às membranas celulares de *H. salinarum* NRC-1.

A participação do produto proteico do gene VNG_RS05855 no processamento de transcritos associados a proteínas de membrana nos faz sugerir que a sua localização pode ser restrita a determinados espaços celulares, supostamente como uma maneira de controlar a sua atividade e reduzir o acesso a um número elevado de substratos. Por outro lado, a suposta compartimentalização desta RNase poderia estar relacionado com a ausência de fenótipos convincentes na linhagem nocaute pela diminuição nas probabilidades de interagir com amplo

grupo de transcritos.

A compartimentalização das RNases em determinadas estruturas ou regiões celulares como uma maneira que a célula tem de controlar a atividade das RNase é uma ideia aceita na comunidade científica (Deutscher, 2015). Vários são os exemplos: a RNase E de *E. coli* e RNase Y de *B. subtilis* estão associadas às membranas celulares internas garantindo a proteção dos RNAs (Campos e Jacobs-Wagner, 2013). A RNase I de *E. coli*, endoribonuclease não específica, está localizada no espaço periplasmático e, como consequência, é incapaz de acessar os RNAs celulares (Deutscher, 2015). Porém não só enzimas endoribonucleases estão associadas a membranas, enzimas exoribonucleolíticas também. Um exemplo é a RNase II de *E. coli*, exoribonuclease inespecífica, na que se comprovou que mutações na sua região N-terminal comprometem a sua interação com a membrana citoplasmática o que implica a perda da sua função e compromete a viabilidade celular (Lu e Taghbalout, 2014). Adicionalmente, outros componentes de processamento e degradação de RNA, incluindo oligoribonuclease, Hfq, poli (A) polimerase, RNase III e pirofosfohidrolase, RppH, têm sido sugeridos como estando presentes em estruturas que envolvem a periferia celular (Taghbalout, Yang e Arluison, 2013).

A localização das RNases longe dos potenciais substratos não descarta a possibilidade da interação com os RNAs para catalisar as reações de processamento ou maturação. Como isso ocorre e os detalhes de como a separação espacial da RNase e do substrato leva, eventualmente, a uma interação produtiva, não é nada clara (Deutscher, 2015). A compartimentalização pode ser uma explicação de como alguma destas enzimas altamente inespecíficas e com atividade geralmente redundantes nas suas linhagens nocaute podem não promover drásticas mudanças fenotípicas (Schilling et al., 2004; Fischer et al., 2012).

5.9 O gene VNG_RS05855 influencia de maneira diferencial o processamento dos transcritos dos distintos replicons de *H. salinarum* NRC-1

Dentro de um organismo a estabilidade dos mRNAs varia consideravelmente, afetando as quantidades de proteínas existentes e influenciando as respostas a estímulos ambientais (Hundt et al., 2007). Os mecanismos de processamento e degradação responsáveis pelas diferenças na estabilidade dos RNA, não só variam entre bactérias e eucariotos, mas dentro do

mesmo domínio mostram marcantes distinções, como a maquinaria de degradação de mRNA de *E. coli* e *B. subtilis* (Condon, 2013) e as meias-vidas dos mRNA de três arqueias, duas espécies *Sulfolobus* e de *Haloferax mediterranei* (Bini et al., 2002; Jäger et al., 2002). Além disso, foi mostrado que as meias-vidas de certos tipos de mRNA são afetadas pela concentração de sal ou a fase de crescimento (Jäger et al., 2002), fornecendo a primeira evidência de que o decaimento do mRNA é um passo importante na regulação gênica em Archaea (Hundt et al., 2007).

Em *H. salinarum* NRC-1 as meias-vidas dos mRNA são consideravelmente mais longas do que as obtidas para *E. coli* e *B. subtilis*, associado ao seu maior tempo de duplicação (Hundt et al., 2007). No entanto, na comparação com organismos com tempo de duplicação semelhante, as meias-vidas de seus mRNA continuam sendo maiores, provavelmente consequência da influência da temperatura no crescimento (Evgenieva-Hackenberg e Klug, 2009), da alta concentração de íons no halófilo, ou da ausência das subunidades do exossomo e de caudas no RNA que levam a uma maior estabilização das moléculas de RNA (Koonin, Wolf e Aravind, 2001; Evgenieva-Hackenberg e Klug, 2009).

A distribuição das meias-vidas dos mRNA de *H. salinarum* NRC-1 diferiu significativamente entre os mRNA codificados nos distintos replicons, sendo a estabilidade dos mRNA codificados no cromossomo maior que as de os plasmídeos (Hundt et al., 2007). Com nossos resultados ao relacionar os tamanhos de cada replicon com os sítios de processamento pós-transcricionais (TPS) neles identificados, tanto pelo estudo das descontinuidades do perfil de cobertura de reads com metodologias implementadas “in house” (Figura 4.26) ou pelos resultados do dRNA-seq analisadas com o software TSSAR stand-alone (Figura 4.33), corroboram este resultado. A densidade de TPS nos transcritos derivados do cromossomo é menor que a densidade dos megaplasmídeos, indicando o processamento diferencial em cada replicon, o que pode resultar na maior estabilidade dos RNAs do cromossomo. Não está claro se as diferenças nas meias-vidas dos mRNA associados aos distintos replicons está relacionada à função ou ao conteúdo GC (Evgenieva-Hackenberg e Klug, 2009), porém o processamento pós-transcricional diferencial para cada replicon de *H. salinarum* NRC-1 parece ser uma característica marcante.

5.10 VNG_RS05855 está envolvido no processamento de transcritos longos

Análises da correlação entre as meias-vidas dos mRNA e o seu comprimento tem mostrado que variam significativamente. Enquanto em *E. coli* e *S. cerevisiae* não foi encontrada uma forte correlação (Bernstein et al., 2002; Wang et al., 2002), a meia-vida do mRNA é negativamente correlacionada com os comprimentos dos transcritos em *Sulfolobus* (Bini et al., 2002). Em *H. salinarum* NRC-1 não foi possível detectar claramente uma correlação significativa entre o comprimento do transcrito e a sua estabilidade. A média das meias-vidas para transcritos de diferentes comprimentos (t^*) ($t^* < 500\text{nt}$; $501\text{nt} < t^* < 1000\text{nt}$; $1001\text{nt} < t^* < 1500\text{nt}$ e $t^* > 1500\text{nt}$) foi similar: 10 ou 11 min (Hundt et al., 2007). Para que transcritos longos tenham meias-vidas similares a transcritos curtos, nós sugerimos que os primeiros sejam alvos de um ativo processamento.

Ao estudar o comprimento dos conjuntos de genes com expressão diferencial, estimados tanto pelo DESeq2 como pelos gráficos de MA-plot, percebemos que mais da metade dos genes com expressão diferencial supera os 600 pb (Figura 4.30), o que indica que independentemente das amostras serem fragmentadas, incluindo ou não a seleção por tamanho de material genético a ser sequenciado, genes de longos comprimentos estão sendo ativamente processados.

5.11 Unidades policistrônicas de *H. salinarum* NRC-1 são diferencialmente processadas na linhagem nocaute do gene VNG_RS05855

Para algumas arqueias e bactérias foi demonstrado a estabilidade diferencial para mRNA individuais de uma mesma unidade policistrônica (Jäger et al., 2002; Dar e Sorek, 2018). No entanto, mediante ensaios de *microrray* (Hundt et al., 2007) mostraram que os segmentos individuais de mRNA policistrônicos em *H. salinarum* NRC-1 tem taxas de decaimento semelhantes. Observações experimentais posteriores feitas por (Koide et al., 2009) com ensaios de *tiling array* mostraram sinais que marcam sítios de início da transcrição (TSS) em genes internos do operons, o que poderia promover níveis diferentes dos mRNA individuais dentro de uma unidade policistrônica. Independentemente dos métodos que utilizamos para estimar os genes que

têm expressão diferencial, as unidades individuais dentro do operon apresentam níveis distintos (Tabelas 4.9, 4.10 e 4.11), sugerindo mais um mecanismo para alterar a estequiometria do operon, o processamento pós-transcricional (Owolabi e Rosen, 1990; Dar e Sorek, 2018).

Um exemplo é observado ao estudar o comportamento das meias-vidas do operon das vesículas de gás em *H. salinarum* NRC-1. Nos estudos realizados por (Hundt et al., 2007) as meias-vidas de todos os mRNA das gvp variaram entre 8 e 9 min. Estudos anteriores mostraram a abundância diferencial das unidades individuais do operon gvpACNO, sendo o transcrito do gvpA mais abundante que os transcritos gvpAC e gvpN (John T. Halladay, Jeffrey G. Jones, 1993). Em *Haloferax mediterranei* o transcrito gvpA tem uma abundância superior aos restantes membros do operon, e a meia-vida dele mostrou-se superior, fato justificado pela maior estabilidade de seu mRNA frente a processamentos pós-transcricionais (Jäger et al., 2002). Dado que as meias-vidas do mRNA de gvpA e os restantes membros do operon das vesículas de gás em *H. salinarum* NRC-1 são semelhantes, as diferenças na sua abundância não foram associadas a processamentos pós-transcricionais e sim a taxas de transcrição diferentes, com uma terminação prematura justamente no fim do gene gvpA, que impede a completa transcrição do operon (Hundt et al., 2007; Evguenieva-Hackenberg e Klug, 2009). Desta maneira, esperaríamos que a abundância deste transcrito em *H. salinarum* NRC-1 não fosse afetada na ausência de ribonucleases que tenham gvpA como possível alvo. No entanto, observamos que na linhagem nocaute do gene VNG_RS05855, a cobertura de reads para gvpA é superior à da linhagem controle, identificando até TPS (Figura 4.35). Isto nos leva a sugerir que exista o processamento pós-transcricional diferencial no transcrito do gene gvpA de *H. salinarum* NRC-1 em relação aos restantes membros do operon gvp, de forma similar a *H. mediterranei* e que apoia o entendimento na expressão diferencial do gene gvpA observado por (John T. Halladay, Jeffrey G. Jones, 1993).

Então, como explicar o fato que as meias-vidas dos transcritos individuais do operons gvp seja semelhante em *H. salinarum* NRC-1 e a sua abundância seja distinta. Uma hipótese é que existam sinais nos transcritos que são reconhecidos por uma ou um conjunto de ribonucleases. Transcritos mais abundantes e com sinais de processamento poderiam ser dinamicamente processados, compartilhando as meias-vidas com aqueles menos abundantes por ter uma trans-

crição e processamento menos ativos, o que poderia levar a melhores respostas adaptativas. Com base nos nossos resultados, não podemos comprovar esta hipótese, no entanto, ao menos do ponto de vista matemático, a regulação mais rápida é a realizada pós-transcricionalmente (Wurtmann et al., 2014), sabe-se do importante papel de sinais estruturais do mRNA nos processos de regulação e que o decaimento diferencial de unidades policistrônicas devem ser direcionados para segmentos específicos do operon, ainda quando os fatores que direcionam os eventos iniciais de clivagem sejam pouco compreendidos (Dar e Sorek, 2018).

5.12 A estruturação do RNA dita o grau de processamento nos transcritos de *H. salinarum* NRC-1

A visão mais comum na relação entre a estruturação do RNA e o processamento pós-transcricional é que a primeira bloqueia a segunda, e como consequência garantem aumentos na estabilidade dos mRNA. A estruturação das extremidades 3' downstream a regiões codificadoras geralmente atrapalham o processamento ribonucleolítico e promovem o decaimento diferencial de transcritos que constituem unidades individuais de sequências policistrônicas de *E. coli*, *B. subtilis* e *M. psychrophilus* R15 (Allenby et al., 2004; Qi et al., 2017; Dar e Sorek, 2018). Em nossos dados, segundo a cobertura de reads dos genes com expressão diferencial, é difícil prever se as estruturas nas extremidades 3' dos transcritos influenciam a atividade do produto proteico do gene VNG_RS05855. Primeiramente, essas regiões não foram mapeadas em nossos experimentos e segundo, observamos nos operons *dpp*, *gvp* e *marR*, que na linhagem controle a suposta atividade exoribonucleolítica do VNG_RS05855 na direção 3'-5' não se paralisa frente a possíveis estruturas. Como consequência, distinguimos a diminuição na cobertura de reads na linhagem controle em relação a sua linhagem nocaute (Figura 4.24 A, Figura 4.25). No entanto, este resultado não nega que eventualmente possam existir estruturas na extremidade 3' e que a interação com o produto proteico do gene VNG_RS05855 influencie o seu processamento.

Por outro lado, conseguimos mapear com o dRNA-seq as extremidades 5' monofosfatadas dos mRNA. A origem dessas extremidades pode ser guiada pela ação exoribonucleolítica na direção 5'-3' ou pela atividade endonucleolítica. A proposta prévia da atividade exoribonucleolítica no sentido 3'-5' do VNG_RS05855 nos encaminhou a sugerir que os enriquecimentos

de extremidades 5' sejam produto da possível ação endoribonucleolítica. Até o momento proteínas com atividade exoribonuclease em ambos sentidos não têm sido descritas. Porém, atividade exo- e endo- ribonucleases para proteínas das famílias das metalo- β -lactamase já foi previamente referida (Dutta e Deutscher, 2009; Dutta e Deutscher, 2010; Dutta, Malhotra e Deutscher, 2012), sendo um exemplo a proteína aCPSF1 (Dominski, Carpousis e Clouet-d'Orval, 2013). Esta proteína tem um motivo característico ausente na sequência proteica da proteína codificada por VNG_RS05855.

Com o dRNA-seq, conseguimos identificar 166 TPS primários associados a atividade do produto proteico do gene VNG_RS05855 nas regiões upstream de genes com expressão diferencial (Figura 4.33 A). Análises das estruturas secundárias das sequências onde os TPS estão inseridos não mostram uma clara conservação, porém é comum encontrá-los dentro de estruturas estáveis (Figura 4.39). A estruturação de regiões 5'-UTR em procariotos tem diferentes funções no processamento do mRNA, podendo tanto evitá-la (Sharp e Bechhofer, 2005) como favorece-la (Qi et al., 2017), em ambos casos o resultado pode ser o mesmo: aumentos na estabilidade dos transcritos.

Se a estabilidade dos transcritos for favorecida pelos processamentos, é esperado que na linhagem nocaute de uma possível RNase a estabilidade dos transcritos diminua. No entanto, não identificamos nenhum TPS nas regiões 5'-UTR dos genes com expressão diminuída na linhagem nocaute. O fato não descarta que o processamento possa ocorrer em regiões na vizinhança do gene com expressão diferencial, internos a outros genes no caso de operons (Figura 4.35). Porém, conseguimos identificar TPS associados a regiões 5'-UTR de 10 genes com expressão aumentada na linhagem nocaute. A identificação de TPS nestas regiões nos sugere que a estruturação não evita o processamento, porque observamos a diminuição na cobertura de reads na linhagem nocaute, talvez a estruturação sirva como sinal para o processamento como foi sugerido por (Qi et al., 2017) em *M. psychrophilus* R15.

A maioria dos TPS associados a atividade do produto proteico do gene VNG_RS05855 correspondem a regiões internas dos genes (Figura 4.33 A). Quando os genes formam parte de operons, a localização interna dos TPS pode ser upstream a genes com expressão diferencial (Figura 4.34 A), downstream (Figura 4.34 B) ou internos ao próprio gene com expressão dife-

rencial (Figura 4.35). Os cortes endoribonucleolíticos nessas regiões podem promover a posterior ação de exoribonucleases, o que levaria a degradação dos transcritos que o constituem. Contudo, em *E. coli* foi mostrado que após do processamento endoribonucleolítico, a formação de estruturas secundárias na extremidade 5' das espécies de mRNA recém geradas são capazes de promover sua estabilidade (Bricker e Belasco, 1999; Dar e Sorek, 2018). As regiões que contêm os TPS internos também são altamente estruturadas (Figura 4.40) e este mecanismo poderia ajudar a explicar ao processamento diferencial para membros de um mesmo operon.

5.13 Processos biológicos afetados na ausência do gene VNG_RS05855

Em *E. coli* estudos de linhagem nocaute da exoribonuclease RNase R, enzima exoribonucleolítica capaz de degradar substratos altamente estruturados, críticas na degradação de tRNA e rRNA defeituosos (Cairrão et al., 2003) afeta a expressão de 202 transcritos relacionados ao transporte, respiração anaeróbica e cadeia de transporte de elétrons (Pobre e Arraiano, 2015). A linhagem nocaute a exoribonuclease RNase II também de *E. coli*, enzima sensível ao processamento de estruturas secundárias, com a capacidade de degradar trechos poliadenilados (Hajnsdorf et al., 1994), afetou a expressão de 187 transcritos relacionados à montagem de flagelo e a motilidade, também processos afetados na ausência da RNase R (Pobre e Arraiano, 2015).

Percebemos que o papel das RNases, seja na degradação ou maturação dos diferentes tipos de RNA é chave para entender as mudanças no metabolismo celular. Elas podem ter diferentes substratos na célula, embora possam também ter sobreposição. Em nossos dados, ao analisar os processos biológicos dos genes com expressão diferencial na ausência de VNG_RS05855, observamos precisamente a diversidade, porém é possível encontrar uma relação entre eles.

Alguns dos genes com expressão diferencial estão relacionados a produção das vesículas de gás. A produção destas estruturas supramoleculares está ligada ao estresse oxidativo (Kaur et al., 2010); este estresse se relaciona o metabolismo do ferro e arsênico por serem estes mediadores das reações de Fenton que junto com os produtos finais da cadeia de transporte de elétrons, também favorecem o aumento de espécies reativas do oxigênio (Hayyan, Hashim e

Alnashef, 2016). Frente ao estresse oxidativo não só se ativam peroxidases e catalases, elementos do metabolismo secundário também são potentes agentes antioxidantes e um exemplo são os sideróforos (Peralta et al., 2016). Sendo conhecido que a produção de vesículas de gás sob as condições de estresses oxidativo é regulada por eventos pós-transcricionais (Kaur et al., 2010), seria esperado que na ausência de uma possível RNase as variações na expressão do conjunto de genes relacionados ao metabolismo de ferro, arsênico, cadeia de transporte de elétrons e agentes antioxidantes tenham algum tipo de relação. Comparando os procedimentos para identificar os genes com expressão diferencial, em muitos deles há sobreposição desses processos biológicos. A interação entre esses processos relatada acima pode ser um exemplo, porém não o único, de como ausência do gene VNG_RS05855 provoque mudanças no metabolismo de *H. salinarum*.

5.14 Atividade ribonuclease da proteína codificada no gene VNG_RS05855

A proteína recombinante de *Haloferax volcanii*, com 66% de identidade de sequência com a codificada no gene VNG_RS05855, teve sua atividade testada frente aos pré-tRNA^{Trp}, pré-tRNA^{Ala} e pré-rRNA5S. A atividade foi avaliada frente a estes substratos porque a sequência compartilha os domínios proteicos da RNase Z. Porém nenhum dos substratos foi clivado pela enzima de *H. volcanii* e os autores concluíram que a falta de atividade era a consequência da ausência do exosite (Fischer et al., 2012). Nas bactérias *Thermotoga maritima*, *E. coli* e *B. subtilis* estudos cristalográficos das proteínas RNase Z evidenciaram diferenças na região do *exosite* com perturbações que iam desde tamanho até a orientação espacial, porém sem comprometer a atividade ribonucleolítica das enzimas nestes organismos (Ishii et al., 2007).

Nós também avaliamos a atividade do produto proteico do gene VNG_RS05855, e para isso a expressamos em *E. coli* e a avaliamos frente a uma maior variedade de substratos: o conjunto de RNA total e o conjunto de RNAs pequenos de *H. salinarum* NRC-1. Em nossos experimentos, estabelecemos a relação 40 µg da proteína para 4 µg de RNA total (relação 10:1), a mesma quantidade de proteína foi testada para 2 µg de RNA pequeno (relação 20:1). Em *H. volcanii* foi avaliada 500 ng da proteína na frente de 2.5 ng substrato (relação 200:1) (Fischer et al., 2012). Em nossos experimentos a relação proteína:substrato foi menor, porém os resultados não deixaram dúvidas da capacidade do produto proteico do gene VNG_RS05855 de processar

o RNA in vitro. Observamos que na presença da proteína o eletroferograma por nano-capilares do RNA total e pequeno é modificado (Figura 4.46, 4.47, 4.48). Até onde temos conhecimento, esta é a primeira descrição da atividade ribonucleolítica para proteínas com domínio metalo- β -lactamase que carecem de exosite na classe Halobacteria.

O perfil do processamento dos pequenos RNA compartilha similaridade com o perfil do processamento das RNase R e RNase A frente ao mesmo tipo de RNA (Figura 4.50 B). A RNase R pode degradar facilmente RNAs altamente estruturados, mas requer uma curta região de fita simples para poder se ligar aos substratos (Pobre e Arraiano, 2015). Entretanto, a RNase A é uma ribonuclease com capacidade de processar substratos de RNA endonucleolíticamente (Moussaoui et al., 1996). Em conjunto, a similaridade do perfil de processamento nos sugere que a proteína VNG_RS05855 seja também capaz de processar substratos estruturados com alguma eficiência, assim como promover de igual maneira rupturas endoribonucleolíticas.

6 Conclusões

A sobrevivência de uma célula depende de sua capacidade de se adaptar rapidamente a mudanças nas condições ambientais, nas quais a biologia do RNA desempenha um papel crítico, pois direciona múltiplos processos mecanicamente distintos (Clouet-d’Orval, 2017). Neste contexto, não é surpreendente que as RNases, reconhecidas como ferramentas de precisão para o RNA, sejam frequentemente essenciais na viabilidade celular (Arraiano et al., 2013). Entender os papéis únicos e compartilhados dessas enzimas na célula é o principal desafio para decifrar as vias de regulação pós-transcricional da expressão gênica. Ainda que muito se conheça sobre a degradação e maturação do RNA em Eukaryas e Bactérias, nas Archaeas o papel destas enzimas continua a ser explorado visto que a identificação e caracterização das RNases estão em estágios iniciais. Fornecer respostas a esses pontos traz perspectivas na compreensão da rota evolutiva destas enzimas (Evguenieva-Hackenberg e Klug, 2009; Clouet-d’Orval, 2017).

Neste cenário, um organismo atrativo para os estudos dos mecanismos de regulação pós-transcricional das Archaeas é o halófilo *Halobacterium salinarum* NRC-1. Deste organismo, além de ser fácil a sua manipulação em laboratório, se dispõem uma enorme quantidade de dados de transcrito, numa ampla variedade de condições, o que permite realizar a análise integradas dos resultados (Koide et al., 2009; Brooks et al., 2014; Wurtmann et al., 2014). No genoma de *H. salinarum* NRC-1 há pelo menos um ortólogo para treze diferentes RNases de linhagens procarióticas e eucarióticas. Para três deles foram construídas linhagens nocautes (Wurtmann et al., 2014). O foco de nossa pesquisa foi a caracterização de um destes três genes, o VNG_RS05855, do qual se carece de estudos funcionais que confirmem a sua atividade. Nós, mediante a integração de abordagens experimentais e computacionais no estudo deste gene chegamos às seguintes conclusões:

- O produto proteico do gene VNG_RS05855 tem função no processamento in vitro dos RNA de *H. salinarum* NRC-1. O processamento ribonucleolítico em proteínas que pertencem a superfamília das metalo- β -lactamases, e que carecem de domínio exosite, não tinha sido demonstrado previamente na classe Halobacteria, tornado-se o primeiro reporte experimental da atividade in vitro deste tipo de proteína.

- O processamento *in vitro* da proteína recombinante, rVNG_RS05855, se caracteriza por não ser de máxima eficiência, visto que não alcança a total degradação dos RNA nas concentrações avaliadas. Frente a pequenos RNA e RNA transportadores a atividade da proteína é mais acentuada e com algum nível de preferência, em virtude da maior intensidade na degradação quando as amostras procedem da linhagem nocaute.
- O gene VNG_RS05855 influencia a o processamento dos transcritos longos e de unidades policistrônicas, ao que contribui o grau de estruturação dos transcritos, o que também poderia ser útil na sinalização de regiões de processamento. Estes processamento se caracteriza por ser diferencial entre os replicons do organismo, atuando menos nos transcritos derivados do cromossomo que nos dos megaplasmídeos e poderia influenciar na estabilidade dos transcritos.
- Sugerimos a inserção do produto proteico do gene VNG_RS05855 na rede de regulação pós-transcricional de *H. salinarum* NRC-1 com destaque na regulação do operon das vesículas de gás e nos transcritos dos diferentes transportadores de membrana. Na regulação das vesículas de gás indicamos um papel no controle dos elementos antisense (asRNA) que poderiam ajustar os níveis de tradução e estabilidade dos membros do operon. Na regulação dos transportadores de membrana, o impacto da sua função é mais heterogêneo, sustentando a complexidade do metabolismo do RNA. Para alguns transcritos, a ausência da proteína aumenta o perfil de cobertura, e em outros, o mesmo efeito é alcançado com a sua presença, sugerindo que afeta a estabilidade. Um exemplo é a falta de processamento nos transcritos de resposta a arsênico na linhagem nocaute, que converte estas células em susceptíveis frente a este metalóide.
- Indicamos a atividade do produto proteico pode ser tanto endo- como exo- ribonucleolítica com direção de processamento 3'-5' com base nas diferenças na perfil de cobertura de reads entre as linhagens controle e nocaute, e a localização dos potenciais sítios de processamento dos transcritos (TPS- transcript processing sites). Esta característica de processamento dual é compartilhada com outros membros da superfamília das metalo- β -lactamases com as que a proteína em estudo não compartilha semelhança em determina-

dos domínios proteicos.

- A ausência do gene VNG_RS05855 não afetou nem a viabilidade nem o crescimento da linhagem nocaute nas condições de referência. A falta de fenótipo pode ser uma consequência da possível localização em membranas da proteína, como tem sido relatado para outras RNases de procariotos e que representa um mecanismo da célula para controlar a atividade destas enzimas.
- Descobrimos sutis diferenças fenotípicas entre as linhagens nocaute e controle frente a determinadas concentrações de agentes estressantes. Em baixas concentrações de Cu^{2+} , Fe^{2+} , Zn^{2+} , Mn^{2+} e Co^{2+} observamos melhorias no crescimento da linhagem nocaute, em H_2O_2 uma ligeira susceptibilidade da linhagem nocaute, de maneira similar frente a mudanças da concentração de sais do meio, porém mais marcada foi a susceptibilidade na frente ao AsO_2^- . As diferenças se concentram nas fases finais da curva, o poderia estar associado a outros os fatores que incrementam a $\text{OD}_{600\text{nm}}$: liberação de vesículas de gás e produção de bacteriorodopsina. Porém, nenhuma destes fatores mostrou diferenças entre as linhagens nos múltiplos ensaios realizados. Em concentrações sub-letais de Cu^{2+} , AsO_2^- e PQT, a sobrevivência da linhagem nocaute se manteve inalterável em relação a linhagem controle, indicando que, no prévio condicionamento ao estresse, VNG_RS05855 não participe nas respostas adaptativas. Também mostramos que o produto proteico deste gene está relacionado com as diferenças no pH diante de transições metabólicas promovidas por mudanças na disponibilidade de nutrientes e luz.
- A expressão do gene VNG_RS05855 pode ser correlacionada a expressão de outros genes que codificam RNases, justificando a ação coordenada de forma cooperativa, redundante ou não específica deste tipo de enzimas. Como consequência, as modificações que podem não ser indispensáveis para atividade de outras RNases, poderiam influenciar com precisão os níveis dos RNA na célula.

As conclusões até aqui relatadas foram obtidas mediante a aplicação de técnicas experimentais aplicadas em larga escala e consequentemente as análises feitas têm este enfoque. Neste sentido, propomos, que com base nos resultados desta pesquisa, novos experimentos sejam pla-

nejados para validar os grupos de transcritos alvos e posições de processamento da atividade ribonucleolítica do produto proteico do gene VNG_RS05855.

7 Anexos

7.1 Curvas de crescimento das linhagens controle (Δ ura3-vermelho) e nocaute (Δ VNG_RS05855-preto) submetidas a diferentes estresses

Curvas de crescimento realizadas em 200 μ L de meio CM sob estresses ocasionados por:

CuSO₄·5H₂O (0.2, 0.4, 0.8, 1.0, 1.5mM)

$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (1.0, 2.0, 4.0, 5.0, 6.0mM)

1'-dimetil-4,4'-bipiridina-dicloreto (PQT) (0.125, 0.250, 0.500, 0.750, 1.00mM)

Curvas de crescimento realizadas em 40 mL de meio CM sob estresses ocasionados por:

MnSO₄·H₂O (0.8, 1.5 mM)

$\text{CoSO}_4 \cdot 7\text{H}_2\text{O}$ (0.2, 0.5 mM)

NaAsO₂ (0.1, 0.2, 0.5 mM)

H₂O₂ (5, 25 mM)

Meio A: NaCl–215g/L, MgSO₄·7H₂O–17.2g/L, Na·Citrate–2.58g/L, KCl–1.72g/L, e peptona 10g/L; Meio B: NaCl–180g/L, MgSO₄ ·7H₂O–14.4g/L, Na·Citrate–2.16g/L, KCl–1.44g/L, e peptone 10g/L

7.2 Genes com expressão diferencial, segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Stranded, identificados como possíveis membros de operons

Tabela 7.1: 139 genes com expressão diferencial (Anexo II, restantes genes), segundo as análises utilizando MA-plot nas bibliotecas preparadas com o protocolo TruSeq-Stranded, identificados como possíveis membros de operons. **Size:** comprimento em pb dos genes. **Member:** gene adjacente na mesma cadeia previsto no mesmo operon. **pOp:** probabilidade estimada de que o par esteja no mesmo operon (valores próximos a 1 ou 0 são previsões confiantes de estar no mesmo operon ou não, enquanto valores próximos a 0.5 são previsões de baixa confiança). **Sep:** distância entre os dois genes, em pares de bases. **MOGScore:** conservação, variando de 0 (não conservado) a 1 (100% conservado). **ExprSim:** correlação de padrões de expressão. **Function:** função do gene com expressão diferencial (Diff. Expres.)

Genes Diff. Expres.	Member	pOp	Sep	MOGScore	ExprSim	Function
VNG_RS00115(up)	VNG_RS13185	0,83	52	1,00	0,862	hypothetical_protein
VNG_RS00835(up)	VNG_RS00830	0,98	-7	0,86	0,845	hypothetical_protein
VNG_RS00905(up)	VNG_RS00900	0,96	0	0,00	0,849	DUF58_domain-containing_protein
VNG_RS01015(up)	VNG_RS01010	0,94	3	0,00	0,882	cytochrome
VNG_RS01100(up)	VNG_RS01095	0,79	28	0,03	0,718	hypothetical_protein
VNG_RS01510(up)	VNG_RS01505	0,93	-15	0,07	NA	4-amino-4-deoxychorismate_lyase
VNG_RS01730(up)	VNG_RS01725	0,77	38	0,00	0,546	dihydrodipicolinate_synthase_family_protein
VNG_RS01940(up)	VNG_RS01945	0,99	-3	0,67	0,813	hypothetical_protein
VNG_RS02010(up)	VNG_RS02005	1,00	-3	0,16	0,426	chromosome_segregation_protein_SMC
VNG_RS02230(up)	VNG_RS02235	0,99	-3	0,66	NA	Na+/H+_antiporter_subunit_G
VNG_RS02230(up)	VNG_RS02225	0,98	-3	0,09	NA	Na+/H+_antiporter_subunit_G
VNG_RS02235(up)	VNG_RS02230	0,99	-3	0,66	NA	cation:proton_antiporter
VNG_RS02235(up)	VNG_RS02240	1,00	-3	0,13	0,914	cation:proton_antiporter
VNG_RS02640(up)	VNG_RS02635	0,97	-3	0,00	0,581	ABC_transporter_permease
VNG_RS02925(up)	VNG_RS02930	0,95	-13	0,76	0,85	AbrB_family_transcriptional_regulator
VNG_RS02930(up)	VNG_RS02925	0,95	-13	0,76	0,85	hypothetical_protein
VNG_RS02930(up)	VNG_RS02935	0,97	-9	0,68	0,926	hypothetical_protein
VNG_RS02935(up)	VNG_RS02930	0,97	-9	0,68	0,926	serine_protein_kinase
VNG_RS03050(up)	VNG_RS03045	0,99	-3	0,25	0,69	hypothetical_protein
VNG_RS03180(up)	VNG_RS03175	0,95	-3	0,00	0,84	spermidine/putrescine_ABC_transporter_substrate-binding_protein
VNG_RS03200(up)	VNG_RS03205	0,98	-3	0,00	0,565	chitinase
VNG_RS03205(up)	VNG_RS03200	0,98	-3	0,00	0,565	hypothetical_protein
VNG_RS03255(up)	VNG_RS03250	1,00	-3	1,00	0,872	4Fe-4S_ferredoxin
VNG_RS03255(up)	VNG_RS03260	1,00	0	1,00	0,871	4Fe-4S_ferredoxin
VNG_RS03385(up)	VNG_RS03380	0,99	0	0,09	0,783	asparagine_synthase
VNG_RS03475(up)	VNG_RS03470	1,00	2	0,31	0,692	DNA_gyrase_subunit_A
VNG_RS04040(up)	VNG_RS04045	0,99	-3	0,22	0,745	GMC_family_oxidoreductase
VNG_RS04440(up)	VNG_RS04435	1,00	-3	0,08	0,923	50S_ribosomal_protein_L13
VNG_RS04460(up)	VNG_RS04455	0,98	2	0,01	0,919	enolase
VNG_RS04460(up)	VNG_RS04465	0,98	-3	0,01	0,957	enolase
VNG_RS04465(up)	VNG_RS04460	0,98	-3	0,01	0,957	30S_ribosomal_protein_S2
VNG_RS04540(up)	VNG_RS04545	0,94	-7	0,03	0,587	methyltransferase
VNG_RS04725(up)	VNG_RS04730	1,00	0	0,11	0,667	urocanate_hydratase
VNG_RS05725(up)	VNG_RS05730	0,98	-3	0,17	0,485	DNA_primase
VNG_RS06070(up)	VNG_RS06075	0,81	27	0,50	NA	hypothetical_protein
VNG_RS06070(up)	VNG_RS06065	0,99	-3	0,90	0,957	hypothetical_protein
VNG_RS06195(up)	VNG_RS06200	0,99	-3	0,19	0,508	sulfate_ABC_transporter_permease
VNG_RS06440(up)	VNG_RS06435	1,00	0	0,50	0,788	rhodopsin
VNG_RS06525(up)	VNG_RS06520	0,96	-7	0,28	0,736	prenyltransferase
VNG_RS06560(up)	VNG_RS06555	1,00	4	0,89	0,947	50S_ribosomal_protein_L4
VNG_RS06560(up)	VNG_RS06565	1,00	-3	0,08	0,939	50S_ribosomal_protein_L4
VNG_RS06615(up)	VNG_RS06620	1,00	-3	0,07	0,962	30S_ribosomal_protein_S4e
VNG_RS06615(up)	VNG_RS06610	1,00	-3	0,48	0,945	30S_ribosomal_protein_S4e
VNG_RS06620(up)	VNG_RS06625	0,99	-3	0,00	0,964	50S_ribosomal_protein_L5
VNG_RS06610	VNG_RS06615	1,00	-3	0,07	0,962	50S_ribosomal_protein_L5
VNG_RS06625(up)	VNG_RS06620	0,99	-3	0,00	0,964	30S_ribosomal_protein_S14
VNG_RS06625(up)	VNG_RS06630	0,99	3	0,00	0,954	30S_ribosomal_protein_S14
VNG_RS06630(up)	VNG_RS06625	0,99	3	0,00	0,954	30S_ribosomal_protein_S8
VNG_RS06630(up)	VNG_RS06635	0,99	4	0,07	0,952	30S_ribosomal_protein_S8
VNG_RS06655(up)	VNG_RS06660	1,00	0	0,48	0,961	30S_ribosomal_protein_S5
VNG_RS06655(up)	VNG_RS06650	0,99	4	0,07	0,905	30S_ribosomal_protein_S5
VNG_RS07065(up)	VNG_RS07070	0,98	-3	0,00	0,586	molybdenum_cofactor_biosynthesis_protein
VNG_RS07165(up)	VNG_RS07160	0,96	0	0,00	0,585	molybdopterin_synthase_sulfur_carrier_subunit
VNG_RS07240(up)	VNG_RS07245	1,00	0	0,25	0,868	spermidine/putrescine_ABC_transporter_permease
VNG_RS07300(up)	VNG_RS07305	0,95	-13	0,06	0,907	aspartate_oxidase
VNG_RS07300(up)	VNG_RS07295	0,99	3	0,02	0,878	aspartate_oxidase
VNG_RS07645(up)	VNG_RS07640	0,98	-7	0,55	0,536	sn-glycerol-3-phosphate_dehydrogenase_subunit_C
VNG_RS07985(up)	VNG_RS07990	0,95	-7	0,30	0,526	dolichol-P-glucose_transferase
VNG_RS08505(up)	VNG_RS08500	0,98	-3	0,88	0,252	hypothetical_protein
VNG_RS08620(up)	VNG_RS08615	0,90	23	0,77	0,365	hypothetical_protein

Tabela 7.2: Continuação_I

VNG_RS08925(up)	VNG_RS08930	0,96	7	0,00	0,848	methylaspartate_mutase
VNG_RS08930(up)	VNG_RS08925	0,96	7	0,00	0,848	glutamate_mutase
VNG_RS08930(up)	VNG_RS08935	1,00	0	0,65	0,783	glutamate_mutase
VNG_RS08935(up)	VNG_RS08930	1,00	0	0,65	0,783	methylaspartate_ammonia-lyase
VNG_RS08945(up)	VNG_RS08940	0,97	-3	0,00	0,891	acetyl-CoA_hydrolase
VNG_RS09230(up)	VNG_RS09235	0,98	27	1,00	0,895	peptide-binding_protein
VNG_RS09235(up)	VNG_RS09240	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS09235(up)	VNG_RS09230	0,98	27	1,00	0,895	peptide_ABC_transporter_permease
VNG_RS09240(up)	VNG_RS09235	1,00	2	0,50	0,835	peptide_ABC_transporter_permease
VNG_RS09275(up)	VNG_RS09270	0,97	-3	0,09	0,378	KEOPS_complex_component
VNG_RS09335(up)	VNG_RS09340	1,00	0	0,56	0,493	ribonucleotide-diphosphate_reductase_subunit_beta
VNG_RS09440(up)	VNG_RS09435	0,95	15	0,40	0,93	hypothetical_protein
VNG_RS09450(up)	VNG_RS09445	0,94	-3	0,50	-0,192	hypothetical_protein
VNG_RS09525(up)	VNG_RS09530	0,82	16	0,00	0,943	hypothetical_protein
VNG_RS09530(up)	VNG_RS09535	0,99	-3	0,01	0,896	argininosuccinate_lyase
VNG_RS09530(up)	VNG_RS09525	0,82	16	0,00	0,943	argininosuccinate_lyase
VNG_RS09535(up)	VNG_RS09530	0,99	-3	0,01	0,896	argininosuccinate_synthase
VNG_RS09915(up)	VNG_RS09910	0,99	2	0,50	0,872	hypothetical_protein
VNG_RS09975(up)	VNG_RS09980	0,97	3	0,00	0,637	ferrichrome_ABC_transporter_permease
VNG_RS10010(up)	VNG_RS10015	0,96	4	0,00	0,659	ferrichrome_ABC_transporter_ATP-binding_protein
VNG_RS10430(up)	VNG_RS10435	0,98	0	0,01	0,62	transcription_elongation_factor_NusA
VNG_RS10435(up)	VNG_RS10430	0,98	0	0,01	0,62	DNA-directed_RNA_polymerase_subunit_A
VNG_RS10730(up)	VNG_RS10735	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_I
VNG_RS10735(up)	VNG_RS10730	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS10745(up)	VNG_RS10740	0,96	3	0,16	NA	ORF_H1831
VNG_RS10750(up)	VNG_RS10755	0,94	9	0,55	NA	ORF_H1818
VNG_RS10775(up)	VNG_RS10780	0,93	4	0,04	NA	ORF_H1786
VNG_RS10780(up)	VNG_RS10775	0,93	4	0,04	NA	ORF_H1776
VNG_RS10900(up)	VNG_RS13160	0,98	-3	0,75	NA	ORF_H0854
VNG_RS11185(up)	VNG_RS11190	0,99	0	0,00	NA	potassium-transporting_ATPase_subunit_B
VNG_RS11185(up)	VNG_RS11180	0,98	2	0,00	NA	potassium-transporting_ATPase_subunit_B
VNG_RS11190(up)	VNG_RS11195	0,88	0	0,00	NA	ATPase
VNG_RS11190(up)	VNG_RS11185	0,99	0	0,00	NA	ATPase
VNG_RS11285(up)	VNG_RS11290	0,98	-3	0,28	NA	aspartate_aminotransferase_family_protein
VNG_RS11290(up)	VNG_RS11285	0,98	-3	0,28	NA	iron_transporter
VNG_RS11290(up)	VNG_RS11295	0,98	-3	0,28	NA	iron_transporter
VNG_RS11300(up)	VNG_RS11305	0,78	26	0,06	NA	lysine_6-monooxygenase
VNG_RS11300(up)	VNG_RS11295	0,97	-3	0,29	NA	lysine_6-monooxygenase
VNG_RS11305(up)	VNG_RS11300	0,78	26	0,06	NA	iron_transporter
VNG_RS11355(up)	VNG_RS11360	0,85	2	0,00	NA	protein_gvpK
VNG_RS11360(up)	VNG_RS11365	0,98	-3	0,50	NA	gas_vesicle_protein_GvpJ
VNG_RS11360(up)	VNG_RS11355	0,85	2	0,00	NA	gas_vesicle_protein_GvpJ
VNG_RS11400(up)	VNG_RS11405	0,97	-3	0,00	NA	gas_vesicle_protein_GvpC
VNG_RS11405(up)	VNG_RS11400	0,97	-3	0,00	NA	ATPase_AAA
VNG_RS11405(up)	VNG_RS11410	0,92	23	0,19	NA	ATPase_AAA
VNG_RS11595(up)	VNG_RS11590	0,88	0	0,00	NA	6-carboxy-5,6,7,8-tetrahydropterin_synthase
VNG_RS12125(up)	VNG_RS12130	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS12430(up)	VNG_RS12425	0,98	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS12450(up)	VNG_RS12445	0,94	9	0,55	NA	ORF_H1813
VNG_RS12960(up)	VNG_RS12955	0,95	-7	0,67	NA	ORF_H1643
VNG_RS13050(up)	VNG_RS13045	0,94	9	0,55	NA	ORF_H1818
VNG_RS13065(up)	VNG_RS13070	0,99	-7	1,00	NA	cytochrome_D_oxidase_subunit_II
VNG_RS00415(down)	VNG_RS00420	0,91	-7	0,00	0,301	RmeM
VNG_RS01010(down)	VNG_RS01015	0,94	3	0,00	0,882	hypothetical_protein
VNG_RS01040(down)	VNG_RS01045	0,97	-7	0,82	0,596	ribonuclease_H
VNG_RS01045(down)	VNG_RS01040	0,97	-7	0,82	0,596	hypothetical_protein
VNG_RS01795(down)	VNG_RS01800	0,95	2	0,00	0,754	transcriptional_regulator
VNG_RS02305(down)	VNG_RS02300	0,94	3	0,00	0,949	hypothetical_protein
VNG_RS02305(down)	VNG_RS02310	0,96	0	0,00	0,755	hypothetical_protein
VNG_RS02370(down)	VNG_RS02365	0,77	119	0,06	0,884	glycosyl_transferase_family_1

Tabela 7.3: Continuação_II

VNG_RS03540(down)	VNG_RS03535	0,96	-10	0,67	0,747	hypothetical_protein
VNG_RS03545(down)	VNG_RS03550	0,80	-16	0,25	0,269	hypothetical_protein
VNG_RS04480(down)	VNG_RS04475	0,98	-3	0,19	0,372	hypothetical_protein
VNG_RS04600(down)	VNG_RS04605	0,96	0	0,00	0,785	hypothetical_protein
VNG_RS05025(down)	VNG_RS05020	1,00	0	1,00	0,852	potassium_transporter_Trk
VNG_RS05400(down)	VNG_RS05405	0,92	-3	0,00	0,583	hypothetical_protein
VNG_RS05470(down)	VNG_RS05465	0,84	26	0,11	0,625	hypothetical_protein
VNG_RS06810(down)	VNG_RS06815	0,78	-3	0,00	0,421	hypothetical_protein
VNG_RS07710(down)	VNG_RS07715	0,77	48	0,42	0,601	hypothetical_protein
VNG_RS07715(down)	VNG_RS07710	0,77	48	0,42	0,601	hypothetical_protein
VNG_RS07810(down)	VNG_RS07815	0,96	5	0,01	0,86	uridylate_kinase
VNG_RS08990(down)	VNG_RS08995	0,78	14	0,00	0,801	hypothetical_protein
VNG_RS10225(down)	VNG_RS10220	0,99	3	0,33	0,889	hypothetical_protein
VNG_RS11590(down)	VNG_RS11595	0,88	0	0,00	NA	radical_SAM_protein
VNG_RS11775(down)	VNG_RS11770	0,97	3	0,43	NA	competence_protein
VNG_RS12445(down)	VNG_RS12450	0,94	9	0,55	NA	ORF_H1818
VNG_RS12550(down)	VNG_RS12545	0,90	6	0,08	NA	ORF_H1617
VNG_RS13370(down)	VNG_RS13375	0,77	14	0,00	0,761	ATP-binding_protein

7.3 OS DADOS DEPOSITADOS EM:

https://drive.google.com/drive/folders/1lmt4Ew77Si_1iVziplYeCxm8wdMB5u4E?usp=sharing

Bibliografia

- Alexa, Adrian, Jörg Rahnenführer e Thomas Lengauer (2006). “Improved scoring of functional groups from gene expression data by decorrelating GO graph structure”. Em: *Bioinformatics* 22.13, pp. 1600–1607. ISSN: 13674803. DOI: [10.1093/bioinformatics/btl1140](https://doi.org/10.1093/bioinformatics/btl1140).
- Allenby, Nicholas E.E. et al. (2004). “Post-transcriptional regulation of the *Bacillus subtilis* pst operon encoding a phosphate-specific ABC transporter”. Em: *Microbiology* 150.8, pp. 2619–2628. ISSN: 13500872. DOI: [10.1099/mic.0.27126-0](https://doi.org/10.1099/mic.0.27126-0).
- Allers, Thorsten e Moshe Mevarech (2005). “Archaeal genetics - The third way”. Em: *Nature Reviews Genetics* 6.1, pp. 58–73. ISSN: 14710056. DOI: [10.1038/nrg1504](https://doi.org/10.1038/nrg1504).
- Almeida, Pereira De et al. (2019). “The Primary Antisense Transcriptome of *Halobacterium salinarum* NRC-1”. Em: *genes* 10, pp. 1–17.
- Ames, James B. et al. (1992). “Resonance Raman study of halorhodopsin photocycle kinetics, chromophore structure, and chloride-pumping mechanism”. Em: *Biochemistry* 31.50, pp. 12546–12554. ISSN: 0006-2960. DOI: [10.1021/bi00165a002](https://doi.org/10.1021/bi00165a002).
- Amman, Fabian et al. (2014). “Supplementary information: “TSSAR: TSS Annotation Regime for dRNA-seq data”. Em: pp. 1–16. URL: <https://static-content.springer.com/esm/art%3A10.1186%2F1471-2105-15-89/MediaObjects/12859%2013%6343%MOESM1%ESM.pdf>.
- Anders, Simon et al. (2010). “Differential expression analysis for sequence count data”. Em: *Genome Biology* 11.10, R106. ISSN: 1465-6906. DOI: [10.1186/gb-2010-11-10-r106](https://doi.org/10.1186/gb-2010-11-10-r106). URL: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2010-11-10-r106>.
- Apiron, David e Andras Miczak (1993). “RNA processing in prokaryotic cells.” Em: *BioEssays* 15, p. 8.
- Arraiano, Cecília Maria et al. (2013). “Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA.” Em: *Biochimica et biophysica acta* 1829.6-7, pp. 491–513. ISSN: 0006-3002. DOI: [10.1016/j.bbagr.2013.03.009](https://doi.org/10.1016/j.bbagr.2013.03.009). URL: <http://www.sciencedirect.com/science/article/pii/S1874939913000540>.
- Babski, Julia et al. (2014). “Small regulatory RNAs in Archaea.” Em: *RNA biology* 11.5, pp. 484–93. ISSN: 1555-8584. DOI: [10.4161/rna.28452](https://doi.org/10.4161/rna.28452). URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4152357&tool=pmcentrez&rendertype=abstract>.
- Bagramyan, K. A. e S. M. Martirosov (1989). “Formation of an ion transport supercomplex in *Escherichia coli* An experimental model of direct transduction of energy”. Em: *FEBS Letters* 246.1-2, pp. 149–152. ISSN: 00145793. DOI: [10.1016/0014-5793\(89\)80272-8](https://doi.org/10.1016/0014-5793(89)80272-8).
- Baliga, Nitin S et al. (2002). “Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach”. Em: *PNAS* 99.23, pp. 1–6. URL: [papers2://publication/uuid/5132640B-4D91-4CD9-83C3-5CD810B31FC7](https://pubs2://publication/uuid/5132640B-4D91-4CD9-83C3-5CD810B31FC7).
- Bandyra, Katarzyna J., Joanna M. Wandzik e Ben F. Luisi (2018). “Substrate Recognition and Autoinhibition in the Central Ribonuclease RNase E”. Em: *Molecular Cell* 72.2, 275–285.e4. ISSN: 10974164. DOI: [10.1016/j.molcel.2018.08.039](https://doi.org/10.1016/j.molcel.2018.08.039). URL: <https://doi.org/10.1016/j.molcel.2018.08.039>.

- Barra, Williams Fernandes et al. (2017). "GEJ cancers: gastric or esophageal tumors? searching for the answer according to molecular identity". Em: *Oncotarget* 8.61, pp. 104286–104294. DOI: [10.18632/oncotarget.22216](https://doi.org/10.18632/oncotarget.22216).
- Barrett, Tanya et al. (2013). "NCBI GEO: Archive for functional genomics data sets - Update". Em: *Nucleic Acids Research* 41.D1, pp. 991–995. ISSN: 03051048. DOI: [10.1093/nar/gks1193](https://doi.org/10.1093/nar/gks1193).
- Bastidas, Oscar (2011). "Conteo Celular con Hematocitómetro". Em: *Technical Note-Neubauer Chamber Cell Counting*, pp. 1–6.
- Bay, Denice C., Kenton L. Rommens e Raymond J. Turner (2008). "Small multidrug resistance proteins: A multidrug transporter family that continues to grow". Em: *Biochimica et Biophysica Acta - Biomembranes* 1778.9, pp. 1814–1838. ISSN: 00052736. DOI: [10.1016/j.bbamem.2007.08.015](https://doi.org/10.1016/j.bbamem.2007.08.015).
- Bebrone, Carine (2007). "Metallo- β -lactamases (classification, activity, genetic organization, structure, zinc coordination) and their superfamily". Em: *Biochemical Pharmacology* 74.12, pp. 1686–1701. ISSN: 00062952. DOI: [10.1016/j.bcp.2007.05.021](https://doi.org/10.1016/j.bcp.2007.05.021).
- Beer, Karlyn D, Elisabeth J Wurtmann e Nitin S Baliga (2014). "Model Organisms Retain an Ecological Memory of Complex Ecologically Relevant Environmental Variation". Em: 80.6, pp. 1821–1831. DOI: [10.1128/AEM.03280-13](https://doi.org/10.1128/AEM.03280-13).
- Bergkessel, Megan, David W. Basta e Dianne K. Newman (2016). "The physiology of growth arrest: Uniting molecular and environmental microbiology". Em: *Nature Reviews Microbiology* 14.9, pp. 549–562. ISSN: 17401534. DOI: [10.1038/nrmicro.2016.107](https://doi.org/10.1038/nrmicro.2016.107). URL: <http://dx.doi.org/10.1038/nrmicro.2016.107>.
- Bernstein, J. A. et al. (2002). "Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays". Em: *Proceedings of the National Academy of Sciences* 99.15, pp. 9697–9702. ISSN: 0027-8424. DOI: [10.1073/pnas.112318199](https://doi.org/10.1073/pnas.112318199). URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.112318199>.
- Bini, Elisabetta et al. (2002). "Stability of mRNA in the hyperthermophilic archaeon Sulfolobus solfataricus." Em: *RNA (New York, N.Y.)* 8.9, pp. 1129–36. ISSN: 1355-8382. DOI: [10.1017/S1355838202021052](https://doi.org/10.1017/S1355838202021052). URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1370327&tool=pmcentrez&rendertype=abstract>.
- Blazewicz, Steven J. et al. (2013). "Evaluating rRNA as an indicator of microbial activity in environmental communities: Limitations and uses". Em: *ISME Journal* 7.11, pp. 2061–2068. ISSN: 17517362. DOI: [10.1038/ismej.2013.102](https://doi.org/10.1038/ismej.2013.102). URL: <http://dx.doi.org/10.1038/ismej.2013.102>.
- Bolger, Anthony M., Marc Lohse e Bjoern Usadel (2014). "Trimmomatic: A flexible trimmer for Illumina sequence data". Em: *Bioinformatics* 30.15, pp. 2114–2120. ISSN: 14602059. DOI: [10.1093/bioinformatics/btu170](https://doi.org/10.1093/bioinformatics/btu170).
- Boneau, C. Alan (1960). "THE EFFECTS OF VIOLATIONS OF ASSUMPTIONS UNDERLYING THE t TEST". Em: *Psychological Bulletin* 57.1.
- Brantl, Sabine (2002). "Antisense-RNA regulation and RNA interference". Em: *Biochimica et Biophysica Acta - Gene Structure and Expression* 1575.1-3, pp. 15–25. ISSN: 01674781. DOI: [10.1016/S0167-4781\(02\)00280-4](https://doi.org/10.1016/S0167-4781(02)00280-4).
- Bricker, Angela L. e Joel G. Belasco (1999). "Importance of a 5' stem-loop for longevity of papA mRNA in Escherichia coli". Em: *Journal of Bacteriology* 181.11, pp. 3587–3590. ISSN: 00219193.

- Brooks, A. N. et al. (2014). “A system-level model for the microbial regulatory genome”. Em: *Molecular Systems Biology* 10.7, pp. 740–740. ISSN: 1744-4292. DOI: [10.15252/msb.20145160](https://doi.org/10.15252/msb.20145160). arXiv: [1701.00732](https://arxiv.org/abs/1701.00732). URL: <http://msb.embopress.org/cgi/doi/10.15252/msb.20145160>.
- Busch, Courtney R. e Jocelyne DiRuggiero (2010). “MutS and MutL are dispensable for maintenance of the genomic mutation rate in the halophilic archaeon *Halobacterium salinarum* NRC-1”. Em: *PLoS ONE* 5.2, pp. 1–8. ISSN: 19326203. DOI: [10.1371/journal.pone.0009045](https://doi.org/10.1371/journal.pone.0009045).
- Büttner, Katharina, Katja Wenig e Karl Peter Hopfner (2005). “Structural framework for the mechanism of archaeal exosomes in RNA processing”. Em: *Molecular Cell* 20.3, pp. 461–471. ISSN: 10972765. DOI: [10.1016/j.molcel.2005.10.018](https://doi.org/10.1016/j.molcel.2005.10.018).
- Cairrão, Fátima et al. (2003). “Cold shock induction of RNase R and its role in the maturation of the quality control mediator SsrA/tmRNA”. Em: *Molecular Microbiology* 50.4, pp. 1349–1360. ISSN: 0950382X. DOI: [10.1046/j.1365-2958.2003.03766.x](https://doi.org/10.1046/j.1365-2958.2003.03766.x).
- Campos, Manuel e Christine Jacobs-Wagner (2013). “Cellular organization of the transfer of genetic information”. Em: *Current Opinion in Microbiology* 16.2, pp. 171–176. ISSN: 13695274. DOI: [10.1016/j.mib.2013.01.007](https://doi.org/10.1016/j.mib.2013.01.007). URL: <http://dx.doi.org/10.1016/j.mib.2013.01.007>.
- Carol M. Schiller, Bruce A. Fowler e James S. Woods (1977). “Effects of Arsenic on Pyruvate Dehydrogenase Activation”. Em: *Environmental Health Perspective* 19.August, pp. 1–2.
- Carthew, Richard W. and Sontheimer, Erik J (2009). “Origins and Mechanisms of miRNAs and siRNAs”. Em: *Cell* 136.4, pp. 642–655. ISSN: 00928674. DOI: [10.1016/j.cell.2009.01.035](https://doi.org/10.1016/j.cell.2009.01.035). [Origins](https://doi.org/10.1016/j.cell.2009.01.035).
- Casimiro-Soriguer, Carlos S., Antonio Muñoz-Mérida e Antonio J. Pérez-Pulido (2017). “Sma3s: A universal tool for easy functional annotation of proteomes and transcriptomes”. Em: *Proteomics* 17.12, pp. 10–13. ISSN: 16159861. DOI: [10.1002/pmic.201700071](https://doi.org/10.1002/pmic.201700071).
- Cavicchioli, Ricardo (2011). “Archaea - Timeline of the third domain”. Em: *Nature Reviews Microbiology* 9.1, pp. 51–61. ISSN: 17401526. DOI: [10.1038/nrmicro2482](https://doi.org/10.1038/nrmicro2482). URL: <http://dx.doi.org/10.1038/nrmicro2482>.
- Cech, Thomas R. e Joan A. Steitz (2014). “The noncoding RNA revolution - Trashing old rules to forge new ones”. Em: *Cell* 157.1, pp. 77–94. ISSN: 10974172. DOI: [10.1016/j.cell.2014.03.008](https://doi.org/10.1016/j.cell.2014.03.008). URL: <http://dx.doi.org/10.1016/j.cell.2014.03.008>.
- Chavarria, Nikita E. et al. (2014). “Archaeal Tuc1/Ncs6 homolog required for wobble uridine tRNA thiolation is associated with ubiquitin-proteasome, translation, and RNA processing system homologs”. Em: *PLoS ONE* 9.6. ISSN: 19326203. DOI: [10.1371/journal.pone.0099104](https://doi.org/10.1371/journal.pone.0099104).
- Chen, Xiaowei S., David Penny e Lesley J. Collins (2011). “Characterization of RNase MRP RNA and novel snoRNAs from *Giardia intestinalis* and *Trichomonas vaginalis*”. Em: *BMC Genomics* 12.1, p. 550. ISSN: 14712164. DOI: [10.1186/1471-2164-12-550](https://doi.org/10.1186/1471-2164-12-550). URL: <http://www.biomedcentral.com/1471-2164/12/550>.
- Clouet-d’Orval, Beatrice (2017). *RNA Metabolism and Gene Expression in Archaea*. Ed. por Beatrice Clouet-d’Orval. 1st ed. 20. Cham, Switzerland: Springer International Publishing, pp. 118–217. ISBN: 9783319657943. DOI: [10.1007/978-3-319-65795-0](https://doi.org/10.1007/978-3-319-65795-0).
- Clouet-D’Orval, Béatrice et al. (2010). “Euryarchaeal β -CASP proteins with homology to bacterial RNase J have 5â² – to3² – exoribonuclease activity”. Em: *Journal of Biological Chemistry* 285.23, pp. 17574–17583. ISSN: 00219258. DOI: [10.1074/jbc.M109.095117](https://doi.org/10.1074/jbc.M109.095117).

- Condon, Ciaran (2013). “RNA Processing and Degradation in *Bacillus subtilis*”. Em: *Microbiology and Molecular Biology Reviews* 67.2, pp. 157–174. ISSN: 0950382X. DOI: [10.1128/MMBR.67.2.157](https://doi.org/10.1128/MMBR.67.2.157). URL: file:///Users/davidmasuda/{_}MASTER/DATA/SENTE/DLMMaster.sente6lib/Contents/Attachments/AB/AB450F99-A19D-44FA-BC27-0719D8562095/122974E6-F1E0-4931-94E0-CC855FF32117.pdf.
- Crooks, G et al. (2004). “WebLogo: A Sequence Logo Generator”. Em: *Genome Res* 14, pp. 1188–1190. ISSN: 1088-9051. DOI: [10.1101/gr.849004.1](https://doi.org/10.1101/gr.849004.1). URL: <ftp://ftp.ncbi.nih.gov/genomes/Bacteria>.
- Danan, Miri et al. (2012). “Transcriptome-wide discovery of circular RNAs in Archaea”. Em: *Nucleic Acids Research* 40.7, pp. 3131–3142. ISSN: 03051048. DOI: [10.1093/nar/gkr1009](https://doi.org/10.1093/nar/gkr1009).
- Daniel M. Sullivan, Benjamin G. Bobay, Douglas J. Kojetin, Richele J. Thompson, Mark Rance, Mark A. Strauch e John Cavanagh (2009). “Insights into the nature of DNA binding of AbrB-like transcription factors”. Em: *Cytometry* 73.11, pp. 1043–1049. ISSN: 0736-6205. DOI: [10.1002/cyto.a.20594](https://doi.org/10.1002/cyto.a.20594). Use. arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Daou-Chabo, Roula e Ciarán Condon (2009). “RNase J1 endonuclease activity as a probe of RNA secondary structure”. Em: *Rna* 15.7, pp. 1417–1425. ISSN: 13558382. DOI: [10.1261/rna.1574309](https://doi.org/10.1261/rna.1574309).
- Dar, Daniel e Rotem Sorek (2018). “Extensive reshaping of bacterial operons by programmed mRNA decay”. Em: *PLoS Genetics* 14.4, pp. 1–19. ISSN: 15537404. DOI: [10.1371/journal.pgen.1007354](https://doi.org/10.1371/journal.pgen.1007354).
- Dar, Daniel et al. (2016). “Widespread formation of alternative 3’ UTR isoforms via transcription termination in archaea”. Em: *Nature Microbiology* 1.10, pp. 1–9. ISSN: 20585276. DOI: [10.1038/nmicrobiol.2016.143](https://doi.org/10.1038/nmicrobiol.2016.143). URL: <http://dx.doi.org/10.1038/nmicrobiol.2016.143>.
- DasSarma, S, Priya Arora (1992). “Genetic analysis of the gas vesicle gene cluster in haloarchaea”. Em: *FEMS Microbiology Letters* 153.1, pp. 1–10. ISSN: 03781097. DOI: [10.1016/s0378-1097\(97\)00238-3](https://doi.org/10.1016/s0378-1097(97)00238-3).
- DasSarma, Priya et al. (2012). “Genome-wide responses of the model archaeon *Halobacterium* sp. strain NRC-1 to oxygen limitation”. Em: *Journal of Bacteriology* 194.20, pp. 5530–5537. ISSN: 00219193. DOI: [10.1128/JB.01153-12](https://doi.org/10.1128/JB.01153-12).
- DasSarma, S et al. (2006). “Post-genomics of the model haloarchaeon *Halobacterium* sp. NRC-1”. Em: *Saline Systems* 2, p. 3. ISSN: 1746-1448. DOI: [10.1186/1746-1448-2-3](https://doi.org/10.1186/1746-1448-2-3). URL: [{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447603/pdf/1746-1448-2-3.pdf">http://www.ncbi.nlm.nih.gov/pubmed/16542428{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC1447603/pdf/1746-1448-2-3.pdf](http://www.ncbi.nlm.nih.gov/pubmed/16542428).
- Deutscher, Murray P. (2006). “Degradation of RNA in bacteria: Comparison of mRNA and stable RNA”. Em: *Nucleic Acids Research* 34.2, pp. 659–666. ISSN: 03051048. DOI: [10.1093/nar/gkj472](https://doi.org/10.1093/nar/gkj472).
- (2015). “How bacterial cells keep ribonucleases under control”. Em: *FEMS Microbiology Reviews* 39.3, pp. 350–361. ISSN: 15746976. DOI: [10.1093/femsre/fuv012](https://doi.org/10.1093/femsre/fuv012).
- Dominski, Zbigniew (2007). “Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism.” Em: *Critical reviews in biochemistry and molecular biology* 42.2, pp. 67–93. ISSN: 1040-9238. DOI: [10.1080/10409230701279118](https://doi.org/10.1080/10409230701279118). URL: <http://www.ncbi.nlm.nih.gov/pubmed/17453916>.
- Dominski, Zbigniew, Agamemnon J. Carpousis e B??atrice Clouet-d’Orval (2013). “Emergence of the beta-CASP ribonucleases: Highly conserved and ubiquitous metallo-enzymes involved in messenger RNA maturation and degradation”. Em: *Biochimica et Biophys-*

- sica Acta - Gene Regulatory Mechanisms 1829.6-7, pp. 532–551. ISSN: 18749399. DOI: [10.1016/j.bbagr.2013.01.010](https://doi.org/10.1016/j.bbagr.2013.01.010). URL: <http://dx.doi.org/10.1016/j.bbagr.2013.01.010>.
- Dutta, Tanmay e Murray P Deutscher (2009). “Catalytic Properties of RNase BN/RNase Z from”. Em: *Journal of Biological Chemistry* 284.23, pp. 15425–15431. DOI: [10.1074/jbc.M109.005462](https://doi.org/10.1074/jbc.M109.005462).
- Dutta, Tanmay e Murray P. Deutscher (2010). “Mode of action of RNase BN/RNase Z on tRNA precursors: RNase BN does not remove the CCA sequence from tRNA”. Em: *Journal of Biological Chemistry* 285.30, pp. 22874–22881. ISSN: 00219258. DOI: [10.1074/jbc.M110.141101](https://doi.org/10.1074/jbc.M110.141101).
- Dutta, Tanmay, Arun Malhotra e Murray P. Deutscher (2012). “Exoribonuclease and endoribonuclease activities of RNase BN/RNase Z both function in vivo”. Em: *Journal of Biological Chemistry* 287.42, pp. 35747–35755. ISSN: 00219258. DOI: [10.1074/jbc.M112.407403](https://doi.org/10.1074/jbc.M112.407403).
- Dyall-Smith, M. (2008). “The Halohandbook: protocols for halobacterial genetics”. Em: *Mark Dyall-Smith, Martinsried, Germany March*, pp. 135–144. DOI: [10.13140/RG.2.1.1750.5441](https://doi.org/10.13140/RG.2.1.1750.5441). URL: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:The+Halohandbook\#0>.
- Edgar, Robert C. (2004). “MUSCLE: Multiple sequence alignment with high accuracy and high throughput”. Em: *Nucleic Acids Research* 32.5, pp. 1792–1797. ISSN: 03051048. DOI: [10.1093/nar/gkh340](https://doi.org/10.1093/nar/gkh340).
- Egli, Pradeep S Pallan e Martin (2008). “Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme- dsDNA complex”. Em: *North* 29.10, pp. 1883–1889. ISSN: 08966273. DOI: [10.3174/ajnr.A1256.Functional](https://doi.org/10.3174/ajnr.A1256.Functional). arXiv: [NIHMS150003](https://arxiv.org/abs/150003).
- Eme, Laura et al. (2017). “Archaea and the origin of eukaryotes”. Em: *Nature Reviews Microbiology* 15.12, pp. 711–723. ISSN: 17401534. DOI: [10.1038/nrmicro.2017.133](https://doi.org/10.1038/nrmicro.2017.133).
- Engelhardt, Jan et al. (2010). “RNAclust Documentation”. Em: *Bioinformatics Group, Department of Computer Science, University of Leipzig*, pp. 1–9. URL: <http://www.bioinf.uni-leipzig.de/~kristin/Software/RNAclust/manual.pdf>.
- Epicentre, Illumina (2012). “Ribonuclease R - epicentre”. Em: pp. 1–3.
- Evguenieva-Hackenberg, Elena e Gabriele Klug (2009). *Chapter 7 RNA Degradation in Archaea and Gram-Negative Bacteria Different from Escherichia coli*. 1ª ed. Vol. 85. C. Elsevier Inc., pp. 275–317. ISBN: 1877-1173 (Print). DOI: [10.1016/S0079-6603\(08\)00807-6](https://doi.org/10.1016/S0079-6603(08)00807-6). URL: [http://dx.doi.org/10.1016/S0079-6603\(08\)00807-6](http://dx.doi.org/10.1016/S0079-6603(08)00807-6).
- (2011). *New aspects of RNA processing in prokaryotes*. DOI: [10.1016/j.mib.2011.07.025](https://doi.org/10.1016/j.mib.2011.07.025).
- Facciotti, Marc T et al. (2007). “General transcription factor specified global gene regulation in archaea.” Em: *Proceedings of the National Academy of Sciences of the United States of America* 104, pp. 4630–4635. ISSN: 0027-8424. DOI: [10.1073/pnas.0611663104](https://doi.org/10.1073/pnas.0611663104).
- Facciotti, Marc T. et al. (2010). “Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis”. Em: *BMC Systems Biology* 4, pp. 1–19. ISSN: 17520509. DOI: [10.1186/1752-0509-4-64](https://doi.org/10.1186/1752-0509-4-64).
- Fedorova, Maria S. et al. (2019). “The CIMP-high phenotype is associated with energy metabolism alterations in colon adenocarcinoma”. Em: *BMC Medical Genetics* 20.Suppl 1. ISSN: 14712350. DOI: [10.1186/s12881-019-0771-5](https://doi.org/10.1186/s12881-019-0771-5).
- Fendrihan, Sergiu, Maurizio Musso e Helga Stan-Lotter (2009). “Raman spectroscopy as a potential method for the detection of extremely halophilic archaea embedded in halite in

- terrestrial and possibly extraterrestrial samples”. Em: *Journal of Raman Spectroscopy* 40.12, pp. 1996–2003. ISSN: 03770486. DOI: [10.1002/jrs.2357](https://doi.org/10.1002/jrs.2357).
- Finn, Robert D. et al. (2016). “The Pfam protein families database: Towards a more sustainable future”. Em: *Nucleic Acids Research* 44.D1, pp. D279–D285. ISSN: 13624962. DOI: [10.1093/nar/gkv1344](https://doi.org/10.1093/nar/gkv1344).
- Fischer, Susan et al. (2012). “Assigning a function to a conserved archaeal metallo- β -lactamase from *Haloferax volcanii*”. Em: *Extremophiles* 16.2, pp. 333–343. ISSN: 14310651. DOI: [10.1007/s00792-012-0433-4](https://doi.org/10.1007/s00792-012-0433-4).
- Franke, Isabel et al. (2003). “YfiK from *Escherichia coli* promotes export of O-acetylserine and cysteine”. Em: *Journal of Bacteriology* 185.4, pp. 1161–1166. ISSN: 00219193. DOI: [10.1128/JB.185.4.1161-1166.2003](https://doi.org/10.1128/JB.185.4.1161-1166.2003).
- Fröhlich, Kathrin S., Konrad U. Förstner e Zemer Gitai (2018). “Post-transcriptional gene regulation by an Hfq-independent small RNA in *Caulobacter crescentus*”. Em: *Nucleic acids research* 46.20, pp. 10969–10982. ISSN: 13624962. DOI: [10.1093/nar/gky765](https://doi.org/10.1093/nar/gky765).
- Gebetsberger, Jennifer et al. (2012). “tRNA-derived fragments target the ribosome and function as regulatory non-coding RNA in *Haloferax volcanii*”. Em: *Archaea* 2012, pp. 10–12. ISSN: 14723646. DOI: [10.1155/2012/260909](https://doi.org/10.1155/2012/260909).
- Gejiao Wang, Sean P. Kennedy, Sabeena Fasiludeen, Christopher Rensing e Shiladitya Das-Sarma (2006). “Arsenic Resistance in *Halobacterium* sp. Strain NRC-1 Examined by Using an Improved Gene Knockout System”. Em: *Journal of Bacteriology* 186.10, pp. 3069–3071. DOI: [10.1128/AEM.72.4.3069](https://doi.org/10.1128/AEM.72.4.3069).
- Gelsinger, Diego Rivera e Jocelyne Diruggiero (2018). “The non-coding regulatory RNA revolution in archaea”. Em: *Genes* 9.3. ISSN: 20734425. DOI: [10.3390/genes9030141](https://doi.org/10.3390/genes9030141).
- Gerstein, Mark B. et al. (2014). “Comparative analysis of the transcriptome across distant species”. Em: *Nature* 512.7515, pp. 445–448. ISSN: 14764687. DOI: [10.1038/nature13424](https://doi.org/10.1038/nature13424).
- Glass, Kimberly e Michelle Girvan (2014). “Annotation enrichment analysis: An alternative method for evaluating the functional properties of gene sets”. Em: *Scientific Reports* 4, pp. 1–9. ISSN: 20452322. DOI: [10.1038/srep04191](https://doi.org/10.1038/srep04191). arXiv: [1208.4127](https://arxiv.org/abs/1208.4127).
- Gomes-Filho, José Vicente et al. (2015). “Sense overlapping transcripts in IS 1341 -type transposase genes are functional non-coding RNAs in archaea”. Em: *RNA Biology* 12.5, pp. 490–500. ISSN: 1547-6286. DOI: [10.1080/15476286.2015.1019998](https://doi.org/10.1080/15476286.2015.1019998). URL: [http://www.tandfonline.com/doi/full/10.1080/15476286.2015.1019998](http://www.tandfonline.com/loi/krnb20http://www.tandfonline.com/doi/full/10.1080/15476286.2015.1019998).
- Gonzalez, Orland et al. (2009). “Systems Analysis of Bioenergetics and Growth of the Extreme Halophile *Halobacterium salinarum*”. Em: *PLoS Computational Biology* 5.4. ISSN: 1553734X. DOI: [10.1371/journal.pcbi.1000332](https://doi.org/10.1371/journal.pcbi.1000332).
- Grade, Biochemistry (2012). “RNase V1”. Em: *AmbionLife* 1.4393905, pp. 1–3.
- Gregor, Dagmar e Felicitas Pfeifer (2005). “In vivo analyses of constitutive and regulated promoters in halophilic archaea”. Em: *Microbiology* 151.1, pp. 25–33. ISSN: 13500872. DOI: [10.1099/mic.0.27541-0](https://doi.org/10.1099/mic.0.27541-0).
- Grove, Anne (2017). “Regulation of Metabolic Pathways by MarR Family Transcription Factors”. Em: *Computational and Structural Biotechnology Journal* 15, pp. 366–371. ISSN: 20010370. DOI: [10.1016/j.csbj.2017.06.001](https://doi.org/10.1016/j.csbj.2017.06.001). URL: <http://dx.doi.org/10.1016/j.csbj.2017.06.001>.
- Grüber, Gerhard et al. (2014). “ATP synthases from archaea: The beauty of a molecular motor”. Em: *Biochimica et Biophysica Acta - Bioenergetics* 1837.6, pp. 940–952. ISSN: 18792650. DOI: [10.1016/j.bbabi.2014.03.004](https://doi.org/10.1016/j.bbabi.2014.03.004). URL: <http://dx.doi.org/10.1016/j.bbabi.2014.03.004>.

- Haft, Daniel H. et al. (2013). "TIGRFAMs and genome properties in 2013". Em: *Nucleic Acids Research* 41.D1, pp. 387–395. ISSN: 03051048. DOI: [10.1093/nar/gks1234](https://doi.org/10.1093/nar/gks1234).
- Hajnsdorf, E. et al. (1994). "Roles of RNase E, RNase II and PNPase in the degradation of the rpsO transcripts of Escherichia coli: stabilizing function of RNase II and evidence for efficient degradation in an *ams pnp rnb* mutant." Em: *The EMBO Journal* 13.14, pp. 3368–3377. DOI: [10.1002/j.1460-2075.1994.tb06639.x](https://doi.org/10.1002/j.1460-2075.1994.tb06639.x).
- Hanson, Bryan A. (2014). "ChemoSpec : An R Package for Chemometric Analysis of Spectroscopic Data and Chromatograms (Package version 2.0-2)". Em: pp. 1–41.
- Hartung, Sophia et al. (2010). "Quantitative analysis of processive RNA degradation by the archaeal RNA exosome". Em: *Nucleic Acids Research* 38.15, pp. 5166–5176. ISSN: 13624962. DOI: [10.1093/nar/gkq238](https://doi.org/10.1093/nar/gkq238).
- Hassanshahian, Mehdi e Jafar Mohamadian (2011). "Isolation and characterization of Halobacterium salinarum from saline lakes in Iran". Em: *Jundishapur Journal of Microbiology* 4.SUPPL. 1. ISSN: 20083645.
- Hauke, Jan e Tomasz Kossowski (2011). "Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data". Em: *Quaestiones Geographicae* 30.2, pp. 87–93. ISSN: 0137-477X. DOI: [10.2478/v10117-011-0021-1](https://doi.org/10.2478/v10117-011-0021-1).
- Hayyan, Maan, Mohd Ali Hashim e Inas M. Alnashef (2016). "Superoxide Ion: Generation and Chemical Implications". Em: *Chemical Reviews* 116.5, pp. 3029–3085. ISSN: 15206890. DOI: [10.1021/acs.chemrev.5b00407](https://doi.org/10.1021/acs.chemrev.5b00407).
- Hervé Pagès, Valerie Obenchain, Martin Morgan (2018). "Package 'GenomicAlignments' March". Em:
- Hölzle, Annette et al. (2008). "Maturation of the 5S rRNA 5' end is catalyzed in vitro by the endonuclease tRNase Z in the archaeon H. volcanii." Em: *RNA (New York, N.Y.)* 14.5, pp. 928–937. ISSN: 1355-8382. DOI: [10.1261/rna.933208](https://doi.org/10.1261/rna.933208).
- Hossain, Sk Tofajjen, Arun Malhotra e Murray P. Deutscher (2016). "How RNase R Degrades Structured RNA". Em: *Journal of Biological Chemistry* 291.15, pp. 7877–7887. ISSN: 0021-9258. DOI: [10.1074/jbc.m116.717991](https://doi.org/10.1074/jbc.m116.717991).
- Huang, Da Wei, Brad T. Sherman e Richard A. Lempicki (2009). "Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists". Em: *Nucleic Acids Research* 37.1, pp. 1–13. ISSN: 03051048. DOI: [10.1093/nar/gkn923](https://doi.org/10.1093/nar/gkn923). arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Hundt, Sonja et al. (2007). "Global analysis of mRNA decay in Halobacterium salinarum NRC-1 at single-gene resolution using DNA microarrays". Em: *Journal of Bacteriology* 189.19, pp. 6936–6944. ISSN: 00219193. DOI: [10.1128/JB.00559-07](https://doi.org/10.1128/JB.00559-07).
- Inoue, Keiichi, Takashi Tsukamoto e Yuki Sudo (2014). "Biochimica et Biophysica Acta Molecular and evolutionary aspects of microbial sensory rhodopsins". Em: *BBA - Bioenergetics* 1837.5, pp. 562–577. ISSN: 0005-2728. DOI: [10.1016/j.bbabbio.2013.05.005](https://doi.org/10.1016/j.bbabbio.2013.05.005). URL: <http://dx.doi.org/10.1016/j.bbabbio.2013.05.005>.
- Ishii, Ryohei et al. (2007). "The structure of the flexible arm of Thermotoga maritima tRNase Z differs from those of homologous enzymes". Em: *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 63.8, pp. 637–641. ISSN: 17443091. DOI: [10.1107/S1744309107033623](https://doi.org/10.1107/S1744309107033623).
- Izquierdo, J. M. et al. (1995). *Changing patterns of transcriptional and post-transcriptional control of β -F1-ATPase gene expression during mitochondrial biogenesis in liver*. DOI: [10.1074/jbc.270.17.10342](https://doi.org/10.1074/jbc.270.17.10342).
- Jäger, Andreas et al. (2002). "Individual gvp transcript segments in Haloferax mediterranei exhibit varying half-lives, which are differentially affected by salt concentration and

- growth phase”. Em: *Nucleic Acids Research* 30.24, pp. 5436–5443. ISSN: 03051048. DOI: [10.1093/nar/gkf699](https://doi.org/10.1093/nar/gkf699).
- Jehlička, Jan, Howell G.M. Edwards e Aharon Oren (2014). “Raman spectroscopy of microbial pigments”. Em: *Applied and Environmental Microbiology* 80.11, pp. 3286–3295. ISSN: 10985336. DOI: [10.1128/AEM.00699-14](https://doi.org/10.1128/AEM.00699-14).
- Jehlička, Jan e Aharon Oren (2013). “Raman spectroscopy in halophile research”. Em: *Frontiers in Microbiology* 4.DEC, pp. 1–7. ISSN: 1664302X. DOI: [10.3389/fmicb.2013.00380](https://doi.org/10.3389/fmicb.2013.00380).
- John T. Halladay, Jeffrey G. Jones, Feng Lin and A. Bruce MacDonald (1993). “The rightward gas vesicles operon in *Halobacterium* plasmid pNRC100: Identification of the gvpA and gvpC products by use of antibody probes and genetic analysis of the region downstream of gvpC”. Em: *Bacteriology* 175.3, pp. 684–692.
- Jones, Jeffrey G., Dennis C. Young e Shiladitya DasSarma (1991). “Structure and organization of the gas vesicle gene cluster on the *Halobacterium halobium* plasmid pNRC100”. Em: *Gene* 102.1, pp. 117–122. ISSN: 03781119. DOI: [10.1016/0378-1119\(91\)90549-Q](https://doi.org/10.1016/0378-1119(91)90549-Q).
- Kahles, André, Jonas Behr e Gunnar Rätsch (2015). “MMR: A tool for read multi-mapper resolution”. Em: *Bioinformatics* 32.5, pp. 770–772. ISSN: 14602059. DOI: [10.1093/bioinformatics/btv624](https://doi.org/10.1093/bioinformatics/btv624).
- Kaper, Patricia B. Lodato e James B. (2008). “Post-transcriptional processing of the LEE4 operon in Enterohemorrhagic *Escherichia coli*”. Em: *Mol Microbiol* 23.1, pp. 1–7. ISSN: 15378276. DOI: [10.1038/jid.2014.371](https://doi.org/10.1038/jid.2014.371). arXiv: [NIHMS150003](https://arxiv.org/abs/150003).
- Karr, Elizabeth A. (2014). *Transcription regulation in the third domain*. 1ª ed. Vol. 89. Elsevier Inc., pp. 101–133. ISBN: 0065-2164 (Print)\r0065-2164. DOI: [10.1016/B978-0-12-800259-9.00003-2](https://doi.org/10.1016/B978-0-12-800259-9.00003-2). URL: <http://dx.doi.org/10.1016/B978-0-12-800259-9.00003-2>.
- Kaur, Amardeep et al. (2006). “A systems view of haloarchaeal strategies to withstand stress from transition metals”. Em: *Genome Research* 16.7, pp. 841–854. ISSN: 10889051. DOI: [10.1101/gr.5189606](https://doi.org/10.1101/gr.5189606).
- Kaur, Amardeep et al. (2010). “Coordination of frontline defense mechanisms under severe oxidative stress”. Em: *Molecular Systems Biology* 6. DOI: [10.1038/msb.2010.50](https://doi.org/10.1038/msb.2010.50).
- Kennedy, Sean P et al. (2001). “Understanding the Adaptation of *Halobacterium* Species NRC-1 to Its Extreme Environment through Computational Analysis of Its Genome Sequence Understanding the Adaptation of *Halobacterium* Species NRC-1 to Its Extreme Environment through Computational Analy”. Em: *Genome Research* 11, pp. 1641–1650. DOI: [10.1101/gr.190201](https://doi.org/10.1101/gr.190201).
- Kim, Daehwan, Ben Langmead e Steven L Salzberg (2016). “HISAT: a fast spliced aligner with low memory requirements”. Em: *Nat Methods* 12.4, pp. 357–360. DOI: [10.1038/nmeth.3317](https://doi.org/10.1038/nmeth.3317). HISAT.
- Kira S. Makarova, Michael Y. Galperin e Eugene V. Koonin National (2016). “Comparative genomic analysis of evolutionarily conserved but functionally uncharacterized membrane proteins in archaea: Prediction of novel components of secretion, membrane remodeling and glycosylation systems”. Em: *Bio* 25.3, pp. 289–313. ISSN: 1527-5418. DOI: [10.1007/s11065-015-9294-9](https://doi.org/10.1007/s11065-015-9294-9). Functional. arXiv: [15334406](https://arxiv.org/abs/15334406).
- Klein, Christian et al. (2005). “The membrane proteome of *Halobacterium salinarum*”. Em: *Proteomics* 5.1, pp. 180–197. ISSN: 16159853. DOI: [10.1002/pmic.200400943](https://doi.org/10.1002/pmic.200400943).

- Klimke, William et al. (2009). “The National Center for Biotechnology Information’s Protein Clusters Database”. Em: *Nucleic Acids Research* 37.SUPPL. 1, pp. 216–223. ISSN: 03051048. DOI: [10.1093/nar/gkn734](https://doi.org/10.1093/nar/gkn734).
- Koide, Tie et al. (2009). “Prevalence of transcription promoters within archaeal operons and coding sequences.” Em: *Molecular systems biology* 5.285, p. 285. ISSN: 1744-4292. DOI: [10.1038/msb.2009.42](https://doi.org/10.1038/msb.2009.42). URL: <http://www.ncbi.nlm.nih.gov/pubmed/19536208>.
- Koonin, E. V., Y. I. Wolf e L. Aravind (2001). “Prediction of the archeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach”. Em: *Genome Research* 11.2, pp. 240–252. ISSN: 10889051. DOI: [10.1101/gr.162001](https://doi.org/10.1101/gr.162001).
- Krüger, Kerstin e Felicitas Pfeifer (1996). “Transcript analysis of the c-vac region and differential synthesis of the two regulatory gas vesicle proteins GvpD and GvpE in Halobacterium salinarum PHH4”. Em: *Journal of Bacteriology* 178.14, pp. 4012–4019. ISSN: 00219193. DOI: [10.1128/jb.178.14.4012-4019.1996](https://doi.org/10.1128/jb.178.14.4012-4019.1996).
- Kulshrestha, Archana (2014). “Arsenic-induced abnormalities in glucose metabolism: Biochemical basis and potential therapeutic and nutritional interventions”. Em: *World Journal of Translational Medicine* 3.2, p. 96. ISSN: 2220-6132. DOI: [10.5528/wjtm.v3.i2.96](https://doi.org/10.5528/wjtm.v3.i2.96).
- Kuroda, Teruo e Tomofusa Tsuchiya (2009). “Multidrug efflux transporters in the MATE family”. Em: *Biochimica et Biophysica Acta - Proteins and Proteomics* 1794.5, pp. 763–768. ISSN: 15709639. DOI: [10.1016/j.bbapap.2008.11.012](https://doi.org/10.1016/j.bbapap.2008.11.012).
- L. Aravind and Eugene V. Koonin (1998). “Natural Classification of Ribonucleases”. Em: *Methods in Enzymology* 341, pp. 1–25.
- Lange, Christian et al. (2007). “Genome-wide analysis of growth phase-dependent translational and transcriptional regulation in halophilic archaea”. Em: *BMC Genomics* 8, pp. 1–16. ISSN: 14712164. DOI: [10.1186/1471-2164-8-415](https://doi.org/10.1186/1471-2164-8-415).
- Le Laz, Sébastien, Audrey Le Goaziou e Ghislaine Henneke (2010). “Structure-specific nuclease activities of Pyrococcus abyssi RNase HII”. Em: *Journal of Bacteriology* 192.14, pp. 3689–3698. ISSN: 10985530. DOI: [10.1128/JB.00268-10](https://doi.org/10.1128/JB.00268-10).
- Lee, Choongdeok, Meong il Kim e Minsun Hong (2017). “Structural and functional analysis of BF2549, a PadR-like transcription factor from Bacteroides fragilis”. Em: *Biochemical and Biophysical Research Communications* 483.1, pp. 264–270. ISSN: 10902104. DOI: [10.1016/j.bbrc.2016.12.155](https://doi.org/10.1016/j.bbrc.2016.12.155). URL: <http://dx.doi.org/10.1016/j.bbrc.2016.12.155>.
- Lee, Kangseok, Jonathan A. Bernstein e Stanley N. Cohen (2002). “RNase G complementation of rne null mutation identifies functional interrelationships with RNase E in Escherichia coli”. Em: *Molecular Microbiology* 43.6, pp. 1445–1456. ISSN: 0950382X. DOI: [10.1046/j.1365-2958.2002.02848.x](https://doi.org/10.1046/j.1365-2958.2002.02848.x).
- Letunic, Ivica, Tobias Doerks e Peer Bork (2015). “SMART: Recent updates, new developments and status in 2015”. Em: *Nucleic Acids Research* 43.D1, pp. D257–D260. ISSN: 13624962. DOI: [10.1093/nar/gku949](https://doi.org/10.1093/nar/gku949).
- Leuko, Stefan et al. (2007). “Carotenoid Analysis of Halophilic Archaea by Resonance Raman Spectroscopy”. Em: *Astrobiology* 7.4, pp. 631–643. ISSN: 1531-1074. DOI: [10.1089/ast.2006.0097](https://doi.org/10.1089/ast.2006.0097).
- Levine, Erel et al. (2007). “Quantitative characteristics of gene regulation by small RNA”. Em: *PLoS Biology* 5.9, pp. 1998–2010. ISSN: 15449173. DOI: [10.1371/journal.pbio.0050229](https://doi.org/10.1371/journal.pbio.0050229).

- Lewin, Alex M. e Ian C. Grieve (2006). “Grouping Gene Ontology terms to improve the assessment of gene set enrichment in microarray data”. Em: *BMC Bioinformatics* 7, pp. 1–9. ISSN: 14712105. DOI: [10.1186/1471-2105-7-426](https://doi.org/10.1186/1471-2105-7-426).
- Lo, Fang-Yin et al. (2014). “Evolution of context dependent regulation by expansion of feast/famine regulatory proteins”. Em: *BMC Systems Biology* 8.1, pp. 1–14. DOI: [10.1186/s12918-014-0122-2](https://doi.org/10.1186/s12918-014-0122-2).
- Long, Qinqiang, Kevin Yang e Qinglin Yang (2015). “Regulation of mitochondrial ATP synthase in cardiac pathophysiology.” Em: *American journal of cardiovascular disease* 5.1, pp. 19–32. ISSN: 2160-200X. URL: <http://www.ncbi.nlm.nih.gov/pubmed/26064790>{\%}0A<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4447074>.
- Lu, Feng e Aziz Taghbalout (2014). “ The Escherichia coli major exoribonuclease RNase II is a component of the RNA degradosome ”. Em: *Bioscience Reports* 34.6, pp. 879–891. ISSN: 0144-8463. DOI: [10.1042/bsr20140113](https://doi.org/10.1042/bsr20140113).
- Ma, Miao et al. (2017). “The crystal structure of Trz1, the long form RNase Z from yeast”. Em: *Nucleic Acids Research* 45.10, pp. 6209–6216. ISSN: 13624962. DOI: [10.1093/nar/gkx216](https://doi.org/10.1093/nar/gkx216).
- MacNeill, Stuart A. (2009). *The haloarchaeal chromosome replication machinery*. DOI: [10.1042/BST0370108](https://doi.org/10.1042/BST0370108).
- Maglott, Donna et al. (2011). “Entrez gene: Gene-centered information at NCBI”. Em: *Nucleic Acids Research* 39.SUPPL. 1, pp. 26–31. ISSN: 03051048. DOI: [10.1093/nar/gkq1237](https://doi.org/10.1093/nar/gkq1237).
- Mann, Hagit et al. (2003). “Eukaryotic RNase P: Role of RNA and Protein Subunits of a Primordial Catalytic Ribonucleoprotein in RNA-Based Catalysis”. Em: *Molecular Cell* 12.4, pp. 925–935. ISSN: 10972765. DOI: [10.1016/S1097-2765\(03\)00357-5](https://doi.org/10.1016/S1097-2765(03)00357-5). URL: <http://www.sciencedirect.com/science/article/pii/S1097276503003575>.
- Marchler-Bauer, Aron et al. (2017). “CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures”. Em: *Nucleic Acids Research* 45.D1, pp. D200–D203. ISSN: 13624962. DOI: [10.1093/nar/gkw1129](https://doi.org/10.1093/nar/gkw1129).
- Marshall, Craig P. et al. (2006). “Vibrational spectroscopy of extant and fossil microbes: Relevance for the astrobiological exploration of Mars”. Em: *Vibrational Spectroscopy* 41.2, pp. 182–189. ISSN: 09242031. DOI: [10.1016/j.vibspec.2006.01.008](https://doi.org/10.1016/j.vibspec.2006.01.008).
- Marujo, Paulo E. et al. (2000). “RNase II removes the oligo(A) tails that destabilize the rpsO mRNA of Escherichia coli”. Em: *RNA* 6.8, pp. 1185–1193. ISSN: 13558382. DOI: [10.1017/S135583820000073X](https://doi.org/10.1017/S135583820000073X).
- Matos, Rute G. et al. (2012). “The only exoribonuclease present in *Haloferax volcanii* has an unique response to temperature changes”. Em: *Biochimica et Biophysica Acta - General Subjects* 1820.10, pp. 1543–1552. ISSN: 03044165. DOI: [10.1016/j.bbagen.2012.06.009](https://doi.org/10.1016/j.bbagen.2012.06.009). URL: <http://dx.doi.org/10.1016/j.bbagen.2012.06.009>.
- McKay, Bruce C. (2013). “Post-Transcriptional Regulation of DNA Damage-Responsive Gene Expression”. Em: *Antioxidants & Redox Signaling* 20.4, pp. 640–654. ISSN: 1523-0864. DOI: [10.1089/ars.2013.5523](https://doi.org/10.1089/ars.2013.5523).
- Mehta, Pankaj, Sidhartha Goyal e Ned S. Wingreen (2008). “A quantitative comparison of sRNA-based and protein-based gene regulation”. Em: *Molecular Systems Biology* 4.221. ISSN: 17444292. DOI: [10.1038/msb.2008.58](https://doi.org/10.1038/msb.2008.58). arXiv: [0809.0527](https://arxiv.org/abs/0809.0527).
- Michael I Love, WolfgangHuber e Simon Anders (2014). “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. Em: pp. 1–21. DOI: [10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).

- Mitchell, Peter (1961). “Coupling of phosphorylation to electron and hydrogen transfer”. Em: *Nature* 4787, pp. 144–148.
- Miyakoshi, M., Y. Chao e J. Vogel (2015). “Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA”. Em: *The EMBO Journal* 34.11, pp. 1478–1492. ISSN: 0261-4189. DOI: [10.15252/embj.201490546](https://doi.org/10.15252/embj.201490546).
- Mohanty, Bijoy K. e Sidney R. Kushner (2003). “Genomic analysis in Escherichia coli demonstrates differential roles for polynucleotide phosphorylase and RNase II in mRNA abundance and decay”. Em: *Molecular Microbiology* 50.2, pp. 645–658. ISSN: 0950382X. DOI: [10.1046/j.1365-2958.2003.03724.x](https://doi.org/10.1046/j.1365-2958.2003.03724.x).
- Moissl-Eichinger, Christine et al. (2018). “Archaea Are Interactive Components of Complex Microbiomes”. Em: *Trends in Microbiology* 26.1, pp. 70–85. ISSN: 18784380. DOI: [10.1016/j.tim.2017.07.004](https://doi.org/10.1016/j.tim.2017.07.004). URL: <http://dx.doi.org/10.1016/j.tim.2017.07.004>.
- Moran-Reyna, Aida e James A Coker (2014). “The effects of extremes of pH on the growth and transcriptomic profiles of three haloarchaea”. Em: *F1000Research*, pp. 1–15. ISSN: 2046-1402. DOI: [10.12688/f1000research.4789.2](https://doi.org/10.12688/f1000research.4789.2). URL: <http://f1000research.com/articles/3-168/v2>.
- Moussaoui, Mohamed et al. (1996). “The role of non-catalytic binding subsites in the endonuclease activity of bovine pancreatic ribonuclease A”. Em: *Journal of Biological Chemistry* 271.9, pp. 4687–4692. ISSN: 00219258. DOI: [10.1074/jbc.271.9.4687](https://doi.org/10.1074/jbc.271.9.4687).
- Naureckiene, Saule e Bernt Eric Uhlin (2003). “In vitro analysis of mRNA processing by RNase E in the pap operon of Escherichia coli”. Em: *Molecular Microbiology* 21.1, pp. 55–68. ISSN: 0950-382X. DOI: [10.1046/j.1365-2958.1996.6121101.x](https://doi.org/10.1046/j.1365-2958.1996.6121101.x).
- Ng, Wailap Victor et al. (2000). “Genome sequence of Halobacterium species NRC-1.” Em: *Proceedings of the National Academy of Sciences of the United States of America* 97.22, pp. 12176–12181. ISSN: 0027-8424. DOI: [10.1073/pnas.190337797](https://doi.org/10.1073/pnas.190337797).
- Nieuwlandt, Daniel T., Elizabeth S. Haas e Charles J. Daniels (1991). “The RNA component of RNase P from the archaebacterium haloferax volcanii”. Em: *Journal of Biological Chemistry* 266.9, pp. 5689–5694. ISSN: 00219258.
- Nogués, M Victòria e Claudi M Cuchillo (2001). “Analysis by HPLC of Distributive Activities and the Synthetic (Back) Reaction of Pancreatic-Type Ribonucleases”. Em: *Nuclease Methods and Protocols* 160.C, pp. 015–024. ISSN: 1064-3745 (Print). DOI: [10.1385/1-59259-233-3:015](https://doi.org/10.1385/1-59259-233-3:015). URL: <http://link.springer.com/10.1385/1-59259-233-3:015>.
- Oesterhelt, D. e W. Stoeckenius (1973). “Functions of a New Photoreceptor Membrane”. Em: *Proceedings of the National Academy of Sciences* 70.10, pp. 2853–2857. ISSN: 0027-8424. DOI: [10.1073/pnas.70.10.2853](https://doi.org/10.1073/pnas.70.10.2853). URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.70.10.2853>.
- Oesterhelt, Dieter e Günther Krippahl (1973). “Light inhibition of respiration in Halobacterium halobium”. Em: *FEBS Letters* 36.1, pp. 72–76. ISSN: 00145793. DOI: [10.1016/0014-5793\(73\)80339-4](https://doi.org/10.1016/0014-5793(73)80339-4).
- Ohtani, Naoto et al. (2004). “Identification of the first archaeal Type 1 RNase H gene from Halobacterium sp. NRC-1: Archaeal RNase HI can cleave an RNA-DNA junction”. Em: *Biochemical Journal* 381.3, pp. 795–802. ISSN: 02646021. DOI: [10.1042/BJ20040153](https://doi.org/10.1042/BJ20040153).
- Oliveira, Vanessa E. de et al. (2010). “Carotenes and carotenoids in natural biological samples: A Raman spectroscopic analysis”. Em: *Journal of Raman Spectroscopy* 41.6, pp. 642–650. ISSN: 03770486. DOI: [10.1002/jrs.2493](https://doi.org/10.1002/jrs.2493).

- Omer, Arina D. et al. (2000). "Homologs of small nucleolar RNAs in Archaea". Em: *Science* 288.5465, pp. 517–522. ISSN: 00368075. DOI: [10.1126/science.288.5465.517](https://doi.org/10.1126/science.288.5465.517).
- Orell, Alvaro et al. (2012). "Inorganic polyphosphates in extremophiles and their possible functions". Em: *Extremophiles* 16.4, pp. 573–583. ISSN: 14310651. DOI: [10.1007/s00792-012-0457-9](https://doi.org/10.1007/s00792-012-0457-9).
- Oren, Aharon (2008). "Microbial life at high salt concentrations: Phylogenetic and metabolic diversity". Em: *Saline Systems* 4.1, pp. 1–13. ISSN: 17461448. DOI: [10.1186/1746-1448-4-2](https://doi.org/10.1186/1746-1448-4-2).
- (2012). "The Function of Gas Vesicles in Halophilic Archaea and Bacteria: Theories and Experimental Evidence". Em: *Life* 3.1, pp. 1–20. ISSN: 2075-1729. DOI: [10.3390/life3010001](https://doi.org/10.3390/life3010001). URL: <http://www.mdpi.com/2075-1729/3/1/1/>.
- Owolabi, J. B. e B. P. Rosen (1990). "Differential mRNA stability controls relative gene expression within the plasmid-encoded arsenical resistance operon". Em: *Journal of Bacteriology* 172.5, pp. 2367–2371. ISSN: 00219193. DOI: [10.1128/jb.172.5.2367-2371.1990](https://doi.org/10.1128/jb.172.5.2367-2371.1990).
- Pande, C., J. K. Lanyi e R. H. Callender (1989). "Effects of various anions on the Raman spectrum of halorhodopsin". Em: *Biophysical Journal* 55.3, pp. 425–431. ISSN: 00063495. DOI: [10.1016/S0006-3495\(89\)82836-X](https://doi.org/10.1016/S0006-3495(89)82836-X). URL: [http://dx.doi.org/10.1016/S0006-3495\(89\)82836-X](http://dx.doi.org/10.1016/S0006-3495(89)82836-X).
- Park, Sun Cheol et al. (2017). "Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR". Em: *Nucleic Acids Research* 45.22, pp. 13080–13093. ISSN: 13624962. DOI: [10.1093/nar/gkx1055](https://doi.org/10.1093/nar/gkx1055).
- Peck, Ronald F, Shiladitya DasSarma e Mark P Krebs (2000). "Homologous gene knockout in the archaeon Halobacterium salinarum with ura3 as a counterselectable marker". Em: *Molecular Microbiology* 35.3, pp. 667–676. ISSN: 0950382X. DOI: [10.1046/j.1365-2958.2000.01739.x](https://doi.org/10.1046/j.1365-2958.2000.01739.x).
- Peralta, Daiana R. et al. (2016). "Enterobactin as part of the oxidative stress response repertoire". Em: *PLoS ONE* 11.6, pp. 1–15. ISSN: 19326203. DOI: [10.1371/journal.pone.0157799](https://doi.org/10.1371/journal.pone.0157799).
- Person, B. et al. (2005). "Functional characterization of sensory rhodopsin II from Halobacterium salinarum expressed in Escherichia coli". Em: *FEBS Letters* 579.14, pp. 3147–3151. DOI: [10.1016/j.febslet.2005.05.010](https://doi.org/10.1016/j.febslet.2005.05.010).
- Pfeifer, Felicitas (2012). "Distribution, formation and regulation of gas vesicles". Em: *Nature Reviews Microbiology* 10.10, pp. 705–715. ISSN: 17401526. DOI: [10.1038/nrmicro2834](https://doi.org/10.1038/nrmicro2834). URL: <http://dx.doi.org/10.1038/nrmicro2834>.
- (2015). "Haloarchaea and the formation of gas vesicles." en. Em: *Life (Basel, Switzerland)* 5.1, pp. 385–402. ISSN: 2075-1729. DOI: [10.3390/life5010385](https://doi.org/10.3390/life5010385). URL: <http://www.mdpi.com/2075-1729/5/1/385/htm>.
- Pfeiffer, F. et al. (2008). "Evolution in the laboratory: The genome of Halobacterium salinarum strain R1 compared to that of strain NRC-1". Em: *Genomics* 91.4, pp. 335–346. ISSN: 08887543. DOI: [10.1016/j.ygeno.2008.01.001](https://doi.org/10.1016/j.ygeno.2008.01.001).
- Phu, Van et al. (2009). "Halobacterium salinarum NRC-1 PeptideAtlas: strategies for targeted proteomics". Em: *Proteome* 7.9, pp. 3755–3764. DOI: [10.1021/pr800031f](https://doi.org/10.1021/pr800031f). [Halobacterium](http://pubs.acs.org/doi/10.1021/pr800031f).
- Phung, Duy Khanh et al. (2013). "Archaeal β -CASP ribonucleases of the aCPSF1 family are orthologs of the eukaryal CPSF-73 factor". Em: *Nucleic Acids Research* 41.2, pp. 1091–1103. ISSN: 03051048. DOI: [10.1093/nar/gks1237](https://doi.org/10.1093/nar/gks1237).

- Pobre, Vânia e Cecília M. Arraiano (2015). "Next generation sequencing analysis reveals that the ribonucleases RNase II, RNase R and PNPase affect bacterial motility and biofilm formation in E. coli". Em: *BMC Genomics* 16.1, pp. 1–12. ISSN: 14712164. DOI: [10.1186/s12864-015-1237-6](https://doi.org/10.1186/s12864-015-1237-6).
- Portnoy, Victoria e Gadi Schuster (2006). "RNA polyadenylation and degradation in different Archaea; roles of the exosome and RNase R". Em: *Nucleic Acids Research* 34.20, pp. 5923–5931. ISSN: 03051048. DOI: [10.1093/nar/gkl763](https://doi.org/10.1093/nar/gkl763).
- Pottanat, Neha Dhingra et al. (2019). "Analysis of the Ribonuclease A Superfamily of Antimicrobial Peptides in Patients Undergoing Chronic Peritoneal Dialysis". Em: *Scientific Reports* 9.1, pp. 1–9. ISSN: 20452322. DOI: [10.1038/s41598-019-44219-x](https://doi.org/10.1038/s41598-019-44219-x).
- Price, Morgan N. et al. (2005). "A novel method for accurate operon predictions in all sequenced prokaryotes". Em: *Nucleic Acids Research* 33.3, pp. 880–892. ISSN: 03051048. DOI: [10.1093/nar/gki232](https://doi.org/10.1093/nar/gki232).
- Pruitt, Kim D., Tatiana Tatusova e Donna R. Maglott (2007). "NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins". Em: *Nucleic Acids Research* 35.SUPPL. 1, pp. 61–65. ISSN: 03051048. DOI: [10.1093/nar/gkl842](https://doi.org/10.1093/nar/gkl842).
- Puustinen, Anne et al. (1991). "Properties of the Two Terminal Oxidases of Escherichia coli". Em: *Biochemistry* 30.16, pp. 3936–3942. ISSN: 15204995. DOI: [10.1021/bi00230a019](https://doi.org/10.1021/bi00230a019).
- Qi, Lei et al. (2017). "Genome-wide mRNA processing in methanogenic archaea reveals post-transcriptional regulation of ribosomal protein synthesis". Em: *Nucleic Acids Research* 45.12, pp. 7285–7298. ISSN: 13624962. DOI: [10.1093/nar/gkx454](https://doi.org/10.1093/nar/gkx454).
- Ramírez, Fidel et al. (2014). "DeepTools: A flexible platform for exploring deep-sequencing data". Em: *Nucleic Acids Research* 42.W1, pp. 187–191. ISSN: 13624962. DOI: [10.1093/nar/gku365](https://doi.org/10.1093/nar/gku365).
- Rebekic, A et al. (2015). "Pearson's or spearman's correlation coefficient â which one to use ?" Em: *Poljoprivreda* 21.2, pp. 47–54. ISSN: 13307142. DOI: [10.18047/poljo.21.2.8](https://doi.org/10.18047/poljo.21.2.8).
- Redko, Yulia, Inés Li de Lasier-Gallay e Ciarán Condon (2007). "When all's zed and done: the structure and function of RNase Z in prokaryotes". Em: *Nature reviews Microbiology* 5.4, pp. 278–286. ISSN: 1740-1526. DOI: [10.1038/nrmicro1622](https://doi.org/10.1038/nrmicro1622). URL: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=17363966&5Cnhttp://www.nature.com/nrmicro/journal/v5/n4/full/nrmicro1622.html.
- Reiche, Kristin (2008). "RNAsoup Documentation". Em: *Fraunhofer Institute for Cell Therapy und Immunology, Perlickstr. 1, D-04103 Leipzig, Germany* December 2008, pp. 1–7.
- Ronny Lorenz, Stephan H Bernhart, Christian Höner zu Siederdisen, Hakim Tafer, Christoph Flamm (2011). "ViennaRNA Package 2.0". Em: *Algorithms for molecular biology* 6, pp. 122–128. ISSN: 1748-7188. DOI: [10.1186/1748-7188-6-26](https://doi.org/10.1186/1748-7188-6-26).
- Ruepp, A e J Soppa (1996). "Fermentative arginine degradation in Halobacterium salinarum (formerly Halobacterium halobium): genes, gene products, and transcripts of the arcRACB gene cluster." Em: *Journal of bacteriology* 178.16, pp. 4942–7. ISSN: 0021-9193. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=178278&tool=pmcentrez&rendertype=abstract>.
- Sánchez-Clemente, Rubén et al. (2018). "Study of pH Changes in Media during Bacterial Growth of Several Environmental Strains". Em: *Proceedings* 2.20, p. 1297. DOI: [10.3390/proceedings2201297](https://doi.org/10.3390/proceedings2201297).

- Santangelo, Thomas J. et al. (2009). “Archaeal intrinsic transcription termination in vivo”. Em: *Journal of Bacteriology* 191.22, pp. 7102–7108. ISSN: 00219193. DOI: [10.1128/JB.00982-09](https://doi.org/10.1128/JB.00982-09).
- Saoura, Makenzie et al. (2017). “Substitutions in conserved regions preceding and within the linker affect activity and flexibility of tRNase ZL, the long form of tRNase Z”. Em: *PLoS ONE* 12.10, pp. 1–17. ISSN: 19326203. DOI: [10.1371/journal.pone.0186277](https://doi.org/10.1371/journal.pone.0186277).
- Schäfer, G, M Engelhard e V Müller (1999). “Bioenergetics of the Archaea.” Em: *Microbiology and molecular biology reviews : MMBR* 63.3, pp. 570–620. ISSN: 1092-2172. DOI: [citeulike-article-id:11186509](https://doi.org/citeulike-article-id:11186509). URL: <http://www.ncbi.nlm.nih.gov/pubmed/10477309>{\%}0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC103747.
- Schierling, Karina et al. (2002). “tRNA 3’ end maturation in archaea has eukaryotic features: the RNase Z from *Haloferax volcanii*.” Em: *Journal of molecular biology* 316.4, pp. 895–902. ISSN: 00222836. DOI: [DOI10.1006/jmbi.2001.5395](https://doi.org/10.1006/jmbi.2001.5395).
- Schilling, Oliver et al. (2004). “Characterization of an *Escherichia coli* *elaC* deletion mutant”. Em: *Biochemical and Biophysical Research Communications* 320.4, pp. 1365–1373. ISSN: 0006291X. DOI: [10.1016/j.bbrc.2004.05.227](https://doi.org/10.1016/j.bbrc.2004.05.227).
- Schilling, Oliver et al. (2005). “Exosite modules guide substrate recognition in the ZiPD/ElaC protein family”. Em: *Journal of Biological Chemistry* 280.18, pp. 17857–17862. ISSN: 00219258. DOI: [10.1074/jbc.M500591200](https://doi.org/10.1074/jbc.M500591200).
- Schlosser, Andreas et al. (1993). “NAD⁺ binding to the *Escherichia coli* K⁺ uptake protein TrkA and sequence similarity between TrkA and domains of a family of dehydrogenases suggest a role for NAD⁺ in bacterial transport”. Em: *Molecular Microbiology* 9.3, pp. 533–543. ISSN: 13652958. DOI: [10.1111/j.1365-2958.1993.tb01714.x](https://doi.org/10.1111/j.1365-2958.1993.tb01714.x).
- Schneider, Caroline a, Wayne S Rasband e Kevin W Eliceiri (2012). “NIH Image to ImageJ: 25 years of image analysis”. Em: *Nature Methods* 9.7, pp. 671–675. ISSN: 1548-7091. DOI: [10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089). URL: <http://dx.doi.org/10.1038/nmeth.2089>.
- Schreiner, Madeleine et al. (2015). “Structure of Halorhodopsin from *Halobacterium salinarum* in a new crystal form that imposes little restraint on the E-F loop”. Em: *Journal of Structural Biology* 190.3, pp. 373–378. ISSN: 10958657. DOI: [10.1016/j.jsb.2015.04.010](https://doi.org/10.1016/j.jsb.2015.04.010). URL: <http://dx.doi.org/10.1016/j.jsb.2015.04.010>.
- Sesto, Nina et al. (2013). “The excludon: A new concept in bacterial antisense RNA-mediated gene regulation”. Em: *Nature Reviews Microbiology* 11.2, pp. 75–82. ISSN: 17401526. DOI: [10.1038/nrmicro2934](https://doi.org/10.1038/nrmicro2934).
- Sharma, Cynthia M. e Jörg Vogel (2014). “Differential RNA-seq: The approach behind and the biological insight gained”. Em: *Current Opinion in Microbiology* 19.1, pp. 97–105. ISSN: 18790364. DOI: [10.1016/j.mib.2014.06.010](https://doi.org/10.1016/j.mib.2014.06.010). arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Sharma, Cynthia M. et al. (2007). “A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites”. Em: *Genes and Development* 21.21, pp. 2804–2817. ISSN: 08909369. DOI: [10.1101/gad.447207](https://doi.org/10.1101/gad.447207).
- Sharma, Cynthia M et al. (2010). “The primary transcriptome of the major human pathogen *Helicobacter pylori*.” Em: *Nature* 464.7286, pp. 250–255. ISSN: 0028-0836. DOI: [10.1038/nature08756](https://doi.org/10.1038/nature08756).
- Sharp, Josh S. e David H. Bechhofer (2005). “Effect of 5â² – proximal element on decay of a model mRNA in *B.* Em: *Molecular Microbiology* 57.2, pp. 484–495. ISSN: 0950382X. DOI: [10.1111/j.1365-2958.2005.04683.x](https://doi.org/10.1111/j.1365-2958.2005.04683.x).

- Shimoni, Yishai et al. (2007). “Regulation of gene expression by small non-coding RNAs: A quantitative view”. Em: *Molecular Systems Biology* 3.138, pp. 1–9. ISSN: 17444292. DOI: [10.1038/msb4100181](https://doi.org/10.1038/msb4100181).
- Sigma (2000). “Nuclease S1”. Em: *Sigma, information* 22890.1997, pp. 22890–22890.
- Skellam, J. G. (1946). “The frequency of the difference between two poisson variates belonging to different populations”. Em: *Journal of the Royal Statistical Society* 109.October, p. 296.
- Skulachev, V P (1976). “Bioenergetics of Tomorrow CONVERSION OF LIGHT ENERGY INTO ELECTRIC ENERGY”. Em: *FEBS Letters* 64.1, pp. 23–25.
- Slonczewski, Joan L. et al. (2009). *Cytoplasmic pH Measurement and Homeostasis in Bacteria and Archaea*. Vol. 55. 09. Elsevier, pp. 1–79,317. ISBN: 9780123747907. DOI: [10.1016/S0065-2911\(09\)05501-5](https://doi.org/10.1016/S0065-2911(09)05501-5). URL: [http://dx.doi.org/10.1016/S0065-2911\(09\)05501-5](http://dx.doi.org/10.1016/S0065-2911(09)05501-5).
- Späth, Bettina et al. (2008). “Two archaeal tRNase Z enzymes: Similar but different”. Em: *Archives of Microbiology* 190.3, pp. 301–308. ISSN: 03028933. DOI: [10.1007/s00203-008-0368-4](https://doi.org/10.1007/s00203-008-0368-4).
- Spearman, C. (1904). “The Proof and Measurement of Association between Two Things: The American Journal of Psychology”. Em: *The American journal of psychology* 15.1, pp. 72–101. ISSN: 00029556. DOI: [10.1037/h0065390](https://doi.org/10.1037/h0065390). URL: <http://www.jstor.org/stable/1412159>.
- Srivastava, Pallavee e Meenal Kowshik (2013). “Mechanisms of metal resistance and homeostasis in Haloarchaea”. Em: *Archaea* 2013. ISSN: 14723646. DOI: [10.1155/2013/732864](https://doi.org/10.1155/2013/732864).
- Strauch, M. et al. (2006). “The SpoOA protein of *Bacillus subtilis* is a repressor of the *abrB* gene.” Em: *Proceedings of the National Academy of Sciences* 87.5, pp. 1801–1805. ISSN: 0027-8424. DOI: [10.1073/pnas.87.5.1801](https://doi.org/10.1073/pnas.87.5.1801).
- Tadokoro, Takashi e Shigenori Kanaya (2009). “Ribonuclease H: Molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes”. Em: *FEBS Journal* 276.6, pp. 1482–1493. ISSN: 1742464X. DOI: [10.1111/j.1742-4658.2009.06907.x](https://doi.org/10.1111/j.1742-4658.2009.06907.x).
- Taghbalout, Aziz, Qingfen Yang e Véronique Arluison (2013). “The *Escherichia coli* RNA processing and degradation machinery is compartmentalized within an organized cellular network”. Em: *Biochemical Journal* 458.1, pp. 11–22. ISSN: 0264-6021. DOI: [10.1042/bj20131287](https://doi.org/10.1042/bj20131287).
- Talaue, Cheryl O. et al. (2016). “Model construction and analysis of respiration in *Halobacterium salinarum*”. Em: *PLoS ONE* 11.3, pp. 1–22. ISSN: 19326203. DOI: [10.1371/journal.pone.0151839](https://doi.org/10.1371/journal.pone.0151839).
- Tatusov, R. L. (2001). “The COG database: new developments in phylogenetic classification of proteins from complete genomes”. Em: *Nucleic Acids Research* 29.1, pp. 22–28. ISSN: 13624962. DOI: [10.1093/nar/29.1.22](https://doi.org/10.1093/nar/29.1.22). URL: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/29.1.22>.
- Technologies, Agilent (2005). “Agilent 2100 Bioanalyzer 2100 Expert User’s Guide Technology Licenses”. Em: pp. 1–366.
- (2006). “Agilent RNA 6000 Nano”. Em: 2006 5067, pp. 1–28. URL: <http://www.webcitation.org/6LUOURTFh>.
- (2016). “Agilent Small RNA”. Em: 2016, pp. 1–28.
- Ten-Caten, Felipe et al. (2018). “Internal RNAs overlapping coding sequences can drive the production of alternative proteins in archaea”. Em: *RNA Biology* 15.8, pp. 1119–1132.

- ISSN: 15558584. DOI: [10.1080/15476286.2018.1509661](https://doi.org/10.1080/15476286.2018.1509661). URL: <https://doi.org/10.1080/15476286.2018.1509661>.
- ThermoScientific (2017). “RNase A, DNase and protease-free”. Em: *New York* 2.September, p. 3.
- Tomii, Kentaro e Minoru Kanehisa (1998). “A comparative analysis of ABC transporters in complete microbial genomes”. Em: *Genome Research* 8.10, pp. 1048–1059. ISSN: 10889051. DOI: [10.1101/gr.8.10.1048](https://doi.org/10.1101/gr.8.10.1048).
- Tonner, Peter D. et al. (2017). “Detecting differential growth of microbial populations with Gaussian process regression”. Em: *Genome Research* 27.2, pp. 320–333. ISSN: 15495469. DOI: [10.1101/gr.210286.116](https://doi.org/10.1101/gr.210286.116).
- Touati, Danièle e Sarah Dubrac (2015). “Fur-mediated transcriptional and post-transcriptional regulation of FeSOD expression in *Escherichia coli*”. Em: *Microbiology* 148.1, pp. 147–156. ISSN: 1350-0872. DOI: [10.1099/00221287-148-1-147](https://doi.org/10.1099/00221287-148-1-147).
- Trey Ideker, Vesteynn Thorsson, Jeffrey Ranish (2001). “Integrated genomic and proteomic analyses of a systematically perturbed metabolic network”. Em: *Science* 292.5518, pp. 929–934. ISSN: 0036-8075. DOI: [10.1126/science.292.5518.929](https://doi.org/10.1126/science.292.5518.929).
- Västermark, Åke et al. (2011). “Functional specialization in nucleotide sugar transporters occurred through differentiation of the gene cluster *EamA* (DUF6) before the radiation of Viridiplantae”. Em: *BMC Evolutionary Biology* 11.1, p. 123. ISSN: 14712148. DOI: [10.1186/1471-2148-11-123](https://doi.org/10.1186/1471-2148-11-123). URL: <http://www.biomedcentral.com/1471-2148/11/123>.
- Vauclare, Pierre et al. (2015). “Molecular adaptation and salt stress response of *Halobacterium salinarum* cells revealed by neutron spectroscopy”. Em: *Extremophiles* 19.6, pp. 1099–1107. ISSN: 14334909. DOI: [10.1007/s00792-015-0782-x](https://doi.org/10.1007/s00792-015-0782-x).
- Vitaliy B. Borisova, Robert B. Gennish, James Hempb, Michael I. Verkhovskyc e Aelozersky (2011). “The cytochrome bd respiratory oxygen reductases”. Em: *Biochimica et biophysica acta* 1807.11, pp. 1398–1413. DOI: [10.1016/j.bbabi.2011.06.016](https://doi.org/10.1016/j.bbabi.2011.06.016). The.
- Vogel, Andreas et al. (2005). “The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties”. Em: 386.December, pp. 1253–1264. DOI: [10.1515/BC.2005.142](https://doi.org/10.1515/BC.2005.142).
- Völkel, Sabrina, Sabrina Fröls e Felicitas Pfeifer (2018). “Heavy Metal Ion Stress on *Halobacterium salinarum* R1 Planktonic Cells and Biofilms”. Em: *Frontiers in Microbiology* 9.December, pp. 1–14. DOI: [10.3389/fmicb.2018.03157](https://doi.org/10.3389/fmicb.2018.03157).
- Wagih, Omar (2016). “RWebLogo”. Em: pp. 1–5.
- WAGNER, Gottfried, Rainer HARTMANN e Dieter OESTERHELT (1978). “Potassium Uniport and ATP Synthesis in *Halobacterium halobium*”. Em: *European Journal of Biochemistry* 89.1, pp. 169–179. ISSN: 14321033. DOI: [10.1111/j.1432-1033.1978.tb20909.x](https://doi.org/10.1111/j.1432-1033.1978.tb20909.x).
- Wagner, Steffen e Gabriele Klug (2007). “An archaeal protein with homology to the eukaryotic translation initiation factor 5A shows ribonucleolytic activity”. Em: *Journal of Biological Chemistry* 282.19, pp. 13966–13976. ISSN: 00219258. DOI: [10.1074/jbc.M701166200](https://doi.org/10.1074/jbc.M701166200).
- Wang, Y. et al. (2002). “Precision and functional specificity in mRNA decay”. Em: *Proceedings of the National Academy of Sciences* 99.9, pp. 5860–5865. ISSN: 0027-8424. DOI: [10.1073/pnas.092538799](https://doi.org/10.1073/pnas.092538799).

- Waterhouse, Andrew M. et al. (2009). “Jalview Version 2-A multiple sequence alignment editor and analysis workbench”. Em: *Bioinformatics* 25.9, pp. 1189–1191. ISSN: 13674803. DOI: [10.1093/bioinformatics/btp033](https://doi.org/10.1093/bioinformatics/btp033).
- Webb, Kimberly M et al. (2013). “Effects of intracellular Mn on the radiation resistance of the halophilic archaeon Halobacterium salinarum”. Em: *Extremophiles* 17.3, pp. 485–497. ISSN: 14310651. DOI: [10.1007/s00792-013-0533-9](https://doi.org/10.1007/s00792-013-0533-9).
- Wei, Wenhao et al. (2019). “Identification of key genes involved in the metastasis of clear cell renal cell carcinoma”. Em: *Oncology Letters* 17.5, pp. 4321–4328. ISSN: 17921082. DOI: [10.3892/ol.2019.10130](https://doi.org/10.3892/ol.2019.10130).
- Werner, Finn e Dina Grohmann (2011). “Evolution of multisubunit RNA polymerases in the three domains of life”. Em: *Nature Reviews Microbiology* 9.2, pp. 85–98. ISSN: 17401526. DOI: [10.1038/nrmicro2507](https://doi.org/10.1038/nrmicro2507). URL: <http://dx.doi.org/10.1038/nrmicro2507>.
- Wilcoxon, Frank (1945). “Individual Comparisons by Ranking Methods”. Em: *Biometrics Bulletin* 1, pp. 119–122.
- Wilk, S.S Shapira e M.B. (2005). “An Analysis of Variance Test for Normality (Complete Samples)”. Em: *Biometrika* 52.6, pp. 311–319. ISSN: 00442380.
- Wilkinson, Steven P. e Anne Grove (2006). “Ligand-responsive transcriptional regulation by members of the MarR family of winged helix proteins”. Em: *Current Issues in Molecular Biology* 8.1, pp. 51–62. ISSN: 14673037.
- Will, Sebastian et al. (2012). “LocARNA-P: Accurate Boundary Prediction and Improved Detection of Structured RNAs”. Em: *Rna* 18, pp. 900–914. DOI: [10.1261/rna.029041.111.10](https://doi.org/10.1261/rna.029041.111.10).
- Willers, Imke M. e José M. Cuezva (2011). “Post-transcriptional regulation of the mitochondrial H⁺-ATP synthase: A key regulator of the metabolic phenotype in cancer”. Em: *Biochimica et Biophysica Acta - Bioenergetics* 1807.6, pp. 543–551. ISSN: 00052728. DOI: [10.1016/j.bbabi.2010.10.016](https://doi.org/10.1016/j.bbabi.2010.10.016).
- Woese, C. R., O. Kandler e M. L. Wheelis (1990). “Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya.” Em: *Proceedings of the National Academy of Sciences* 87.12, pp. 4576–4579. ISSN: 0027-8424. DOI: [10.1073/pnas.87.12.4576](https://doi.org/10.1073/pnas.87.12.4576). URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.87.12.4576>.
- Woodson, J D, R F Peck e M P Krebs (2003). “The cobY Gene of the Archaeon Halobacterium sp. Strain NRC-1 Is Required for De Novo Cobamide Synthesis”. Em: 185.1, pp. 311–316. DOI: [10.1128/JB.185.1.311](https://doi.org/10.1128/JB.185.1.311).
- Wurtmann, Elisabeth J. et al. (2014). “An evolutionarily conserved RNase-based mechanism for repression of transcriptional positive autoregulation”. Em: *Molecular Microbiology* 92.2, pp. 369–382. ISSN: 13652958. DOI: [10.1111/mmi.12564](https://doi.org/10.1111/mmi.12564). arXiv: [NIHMS150003](https://arxiv.org/abs/NIHMS150003).
- Xu, Chenggang et al. (2015). “Cellulosome stoichiometry in Clostridium cellulolyticum is regulated by selective RNA processing and stabilization”. Em: *Nature Communications* 6, pp. 1–13. ISSN: 20411723. DOI: [10.1038/ncomms7900](https://doi.org/10.1038/ncomms7900). URL: <http://dx.doi.org/10.1038/ncomms7900>.
- Xu, Jian-zhong, Jun-lan Zhang e Wei-guo Zhang (2018). “Antisense RNA: the new favorite in genetic research”. Em: *Journal of Zhejiang University-SCIENCE B* 19.10, pp. 739–749. ISSN: 1673-1581. DOI: [10.1631/jzus.b1700594](https://doi.org/10.1631/jzus.b1700594).
- Xue Hai, Liang et al. (2005). “A genome-wide analysis of CD and HACA-like small nucleolar RNAs in Trypanosoma brucei reveals a trypanosome-specific pattern of rRNA modification”. Em: *RNA Society* 11, pp. 619–645. DOI: [10.1261/rna.7174805.2](https://doi.org/10.1261/rna.7174805.2).

- Yanjie Chao Lei Li, Dylan Girodat Ben F Luisi (2017). “In Vivo Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways”. Em: pp. 39–51. DOI: [10.1016/j.molcel.2016.11.002](https://doi.org/10.1016/j.molcel.2016.11.002).
- Yao, Andrew I. e Marc T. Facciotti (2011). “Regulatory multidimensionality of gas vesicle biogenesis in *Halobacterium salinarum* NRC-1”. Em: *Archaea* 2011, p. 716456. ISSN: 1472-3654. DOI: [10.1155/2011/716456](https://doi.org/10.1155/2011/716456). URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3202098&tool=pmcentrez&rendertype=abstract>.
- Zaramela, Livia S et al. (2014). “Transcription start site associated RNAs (TSSaRNAs) are ubiquitous in all domains of life.” Em: *PloS one* 9.9, e107680. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0107680](https://doi.org/10.1371/journal.pone.0107680). URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4169567&tool=pmcentrez&rendertype=abstract>.
- Zaramela, Lívia Soares (2015). “Mapeamento de RNAs não-codificantes e suas interações com a chaperona Lsm na archaea *Halobacterium salinarum* Mapeamento de RNAs não-codificantes e suas interações com a chaperona Lsm na archaea *Halobacterium salinarum*”. Em: *Universidade de São Paulo. Faculdade de Medicina de Ribeirão Preto*.
- Zhang, Zong Hong et al. (2014). “A comparative study of techniques for differential expression analysis on RNA-seq data”. Em: *PLoS ONE* 9.8. ISSN: 19326203. DOI: [10.1371/journal.pone.0103207](https://doi.org/10.1371/journal.pone.0103207).
- Zuker, Michael (2003). “Mfold web server for nucleic acid folding and hybridization prediction”. Em: *Nucleic Acids Research* 31.13, pp. 3406–3415. ISSN: 03051048. DOI: [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595).