



SÉRGIO PAULO DEJATO DA ROCHA

IDENTIFICAÇÃO IN SILICO DE PEQUENOS RNAs NÃO CODIFICANTES
EM GENOMA DE PROTEUS MIRABILIS UROPATOGÊNICO

MONOGRAFIA DE ESPECIALIZAÇÃO

LONDRINA

2016

SÉRGIO PAULO DEJATO DA ROCHA

IDENTIFICAÇÃO IN SILICO DE PEQUENOS RNAs NÃO CODIFICANTES
EM GENOMA DE PROTEUS MIRABILIS UROPATOGÊNICO

Monografia apresentada ao Programa de Pós Graduação em Bioinformática da Universidade Tecnológica Federal do Paraná – UTFPR – *Campus* Londrina, como requisito parcial para obtenção do título de Especialista em Bioinformática.

Orientador: Prof.Dr. Alexandre Rossi Paschoal

LONDRINA

2016

SÉRGIO PAULO DEJATO DA ROCHA

IDENTIFICAÇÃO IN SILICO DE PEQUENOS RNAs NÃO CODIFICANTES
EM GENOMA DE PROTEUS MIRABILIS UROPATOGÊNICO

BANCA EXAMINADORA

Orientador: Prof. Dr. Alexandre Rossi Paschoal
Universidade Tecnológica Federal do Paraná- UTFPR

Prof. Dr. Laurival Antônio Villas Bôas
Universidade Estadual de Londrina - UEL

Prof. Dr. Alessandro Botelho Bovo
Universidade Tecnológica Federal do Paraná- UTFPR

Londrina, 22 de agosto de 2016.

Dedico este trabalho à minha esposa
Cristiani e ao meu filho Guilherme
pelo apoio, paciência, confiança, e
amor incondicional.

AGRADECIMENTOS

Ao meu orientador, **Prof. Dr. Alexandre Rossi Paschoal**, pela disposição, paciência e dedicação.

Ao **Prof. Dr. Laurival Villas Boas**, pela generosidade e colaboração.

Ao **Ivan Rodrigo Wolf**, pela generosidade, paciência e imprescindível ajuda nas análises.

Ao **Prof. Dr. Alessandro Botelho Bovo**, pela amizade e organização do curso.

Aos amigos **Alexandre Morey e Juan Sarmiento**, pela amizade e companheirismo nestes anos do curso.

Aos professores do curso de especialização em bioinformática, que sempre tiveram paciência e dedicação em ensinar.

Por fim a toda minha família e amigos que sempre me apoiaram em mais esta conquista

RESUMO

O trato urinário representa um sítio frequente de infecções tanto em pacientes da comunidade como em pacientes internados em unidades hospitalares. A infecção do trato urinário (ITU) caracteriza-se pela multiplicação da bactéria em qualquer parte deste local, seja nos rins, ureteres, bexiga ou uretra. A infecção do trato urinário associada a cateter (ITU-CA) é a infecção nosocomial mais comum dentre todas, e representa mais de 80% das ITU nosocomiais. *Proteus mirabilis*, uma bactéria Gram-negativa, não é um comum causador de ITU em pacientes normais, sendo mais relacionado à ITU complicada, principalmente à ITU-CA. *P. mirabilis* possui dezenas de genomas sequenciados cujo tamanho varia de 3 a 5 Mb. Pequenos RNAs (sRNAs) são sequências curtas de RNA que regulam a expressão gênica. Estas moléculas atuam através do seu pareamento com sequências alvo por meio de complementaridade específica ou parcial de bases. Em procariotos os sRNAs desempenham papéis fundamentais em redes reguladoras de expressão nas respostas à estímulos ambientais, inclusive em bactérias patogênicas. A abordagem computacional é considerada uma das mais eficientes para localização de candidatos a sRNAs em sequências de genomas. Podendo ser dividida de acordo com o método de busca utilizado; dentre os quais, podem ser citados os que procuram estruturas secundárias consensuais; os que efetuam na busca por sinais de transcrição raros; e os que aplicam genômica comparativa e *ab initio*. Diante disso, este trabalho realizou a identificação e caracterização de pequenos RNAs não-codificantes no genoma de *P. mirabilis* por meio de análises de bioinformática (*in silico*). Em suma, os candidatos a sRNA do genoma da cepa *P. mirabilis* HI4320 foram preditos utilizando-se os programas INFERNAL/Rfam e nocoRNAc. Após uma análise manual dos resultados de anotação, foram considerados um total de 29 sRNA. Em geral, os tamanhos variaram de 41 a 505 nucleotídeos, com média de 273 nucleotídeos, sendo que oito tem anotação funcional relacionados com descrito na literatura. Por exemplo, seis já foram descritos em *Escherichia coli* e duas em *Lactococcus lactis*. Em relação as famílias, quatro são cis-regulatórios e quatro são sRNA, sendo que todos se localizam no cromossomo bacteriano. Quanto à função, todos estão envolvidos na regulação da expressão de proteínas de vias metabólicas de procariotos. Portanto, estes resultados são inéditos e podem contribuir no conhecimento da regulação da expressão gênica deste uropatógeno. Para que estas

moléculas deixem de ser somente preditas uma futura validação experimental *in vitro* torna-se necessária bem como a busca pelos seus salvos e o entendimento de suas funções regulatórias.

Palavras-chave: pequenos RNAs, *Proteus mirabilis*, regulação gênica, non-coding RNA, bioinformática.

ABSTRACT

The urinary tract is a frequent site of infections in both community and admitted to hospitals patients. The urinary tract infection (UTI) is characterized by proliferation of bacteria in any part of the urinary tract: kidney, ureters, bladder or urethra. The urinary tract infection associated with catheter (ITU-CA) is the most common nosocomial infection and represents more than 80% of nosocomial UTI. *Proteus mirabilis*, a Gram-negative bacterium, is not a common cause of UTI in normal patients, being more related to complicated ITU, especially the ITU-CA. *P. mirabilis* has dozens of sequenced genomes whose size ranges from 3 to 5 Mb. Small RNAs (sRNAs) are short sequences of RNA that regulate gene expression. These molecules act pairing with the target sequences by specific or partially complementary bases. In prokaryotes the sRNAs play key roles in regulatory networks of expression in response to environmental stimuli, including pathogenic bacteria. The computational approach is considered one of the most efficient for locating new sRNAs. It can be divided according to the search method used; among which may be cited those seeking consensus secondary structures; those who carry out the search for rare transcription signals; and those applying comparative genomics and *ab initio*. Thus, this work constitutes the identification and characterization of small non-coding RNAs in the genome of *P. mirabilis* through *in silico* analysis. The sRNA candidates in the genome of *P. mirabilis* HI4320 strain were predicted using the INFERNAL/Rfam and nocoRNAC programs. In total 29 were predicted sRNA whose sizes ranged from 41 to 505 nucleotides, averaging 273 nucleotides. Of this total only eight could be characterized. Six has been described in *Escherichia coli* and two in *Lactococcus lactis* genomes. Regarding the families, four are cis-reg and four are genes-sRNA and all are located in the bacterial chromosome. Concerning the function, they are all involved in regulating of proteins expression in prokaryotes metabolic pathways. Therefore, these results are unpublished and may contribute to the knowledge of the regulation of gene expression in this uropathogen. For these molecules are not be only predicted a future *in vitro* experimental validation becomes necessary as well as the search for the targets and understanding of its regulatory functions.

Keywords: small RNAs, *Proteus mirabilis*, gene regulation, non-coding RNA, bioinformatic.

LISTA DE ILUSTRAÇÕES

Figura 1- Fluxograma dos procedimentos utilizados para identificar os genes de sRNA no genoma de *P. mirabilis*.....21

Figura 2- Genes dos pequenos RNAs não codificadores preditos para *P. mirabilis* uropatogênico.....22

LISTA DE TABELAS

Tabela 01- Caracterização das sequências de ncRNA pelo nome, tipo, número e regulação na célula bacteriana.....	25
--	----

SUMÁRIO

1 INTRODUÇÃO	12
1.1 OBJETIVOS	13
1.1.1 Objetivo Geral	13
1.1.2 Objetivos Específicos	13
1.2 JUSTIFICATIVA	13
1.3 EMBASAMENTO TEÓRICO	14
2.DESENVOLVIMENTO.....	15
2.Revisão da Literatura.....	15
2.1.1 Infecção do trato urinário (ITU).....	15
2.1.2 <i>Proteus</i> sp	16
2.1.3 Genômica de <i>Proteus mirabilis</i>	17
2.1.4 Pequenos RNAs (sRNAs).....	17
2.1.5 Análise <i>in silico</i> para localização de sRNAs.....	18
2.1.5.1 Infernal.....	18
2.1.5.2 nocoRNAc.....	18
2.1.5.3 Rfam.....	19
2.2 MATERIAL E MÉTODOS.....	19
2.2.1 Sequência de genoma bacteriano.....	20
2.2.2 Anotação.....	20
2.3 RESULTADOS E DISCUSSÃO.....	20
3 CONCLUSÃO.....	27
REFERÊNCIAS.....	28

1- INTRODUÇÃO

O trato urinário representa um sítio frequente de infecções tanto em pacientes da comunidade como em pacientes internados em unidades hospitalares, sendo comuns em todas as idades, e estão entre as causas que mais requerem tratamento ambulatorial (MASSON et al., 2009; RORIZ-FILHO et al., 2010). A infecção do trato urinário (ITU) é referida como qualquer resposta inflamatória urotelial resultante de invasão bacteriana no trato urinário (GONZALEZ; SCHAEFFER, 1999).

A prevalência de bactérias é variável, sendo dependente de fatores como idade, sexo, gravidez, instrumentação e local da aquisição da infecção, se intra ou extra-hospitalar. Porém, de maneira geral, *Escherichia coli* se apresenta como a bactéria mais frequente (RIYUZO et al., 2007; LO et al., 2010; DARZE et al., 2011; MARQUES et al., 2012; LO et al., 2013; TASBAKAN et al., 2013; MORAES et al., 2014). As demais bactérias apresentam uma frequência variável sendo, as espécies de *Proteus*, muitas vezes citadas entre as maiores responsáveis pelas frequências de infecção (HANNAN, 2012; MARQUES et al., 2012; KHAWCHAROENPORN et al., 2013; MODY & JUTHANI-MEHTA, 2014), comuns especialmente em pacientes com dispositivos urológicos internos e/ou anormalidades anatômicas do trato urinário (NICOLLE, 2005; CHEN et al., 2012).

Um RNA não-codificante (ncRNA) é qualquer molécula de RNA que não é traduzida em proteína e o termo pequeno RNA (*small RNA*, *sRNA*) é ainda usado para bactéria. Os sRNAs foram descritos inicialmente em procaríotos, sendo que o primeiro foi descrito em *E. coli* (MIZUNO, et al., 1984). Os sRNAs podem ter funções estruturais, catalíticas ou regulatórias (EDDY, 2001). Estas moléculas atuam através do seu pareamento com sequências alvo por meio de complementaridade específica ou parcial de bases (GOTTESMAN e STORZ, 2011). Assim, conseguem inibir a transcrição ao bloquear sítios de ligação aos ribossomos, e induzir a transcrição de mRNAs, impedindo a formação de estruturas secundárias em sua região 5' não traduzida (5' UTR) (STORZ et al., 2011). A abordagem computacional via bioinformática é considerada uma das mais eficientes para localização de novos sRNAs em genomas (EDDY, 2001).

Proteus mirabilis possui diversos fatores de virulência que podem ser regulados por pequenos RNAs não codificadores. Até o momento não existem descrições destes RNAs de *P. mirabilis* na literatura. Portanto a descrição destes

ácidos nucléicos torna-se importante fonte de conhecimento para futuros estudos averiguarem a participação destas moléculas na regulação da expressão gênica de fatores de virulência da uropatogênese de *P. mirabilis*.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Identificar e caracterizar pequenos RNAs não-codificantes no genoma de *P. mirabilis* por meio de análises de bioinformática (*in silico*).

1.1.2 Objetivos específicos

A seguir são apresentados os objetivos específicos para a realização deste trabalho:

- Identificar pequenos RNAs em *P. mirabilis*.
- Aplicar abordagens híbridas *in silico* para caracterização dos RNAs.
- Validar os resultados por meio da interpretação e comparação com a literatura.

1.2 JUSTIFICATIVA

A infecção do trato urinário é umas das doenças infecciosas que atinge milhões de pessoas por mês no mundo. *P. mirabilis* é um dos principais agentes causais desta doença. Sabe-se que os pequenos RNAs não-codificadores podem regular a expressão gênica em procariotos. *P. mirabilis* possui diversos fatores de virulência que podem ser regulados por pequenos RNAs não codificadores. Até a data de 25/06/2016 não existem descrições destes RNAs de *P. mirabilis* na literatura. Por outro lado, abordagens de bioinformática estão num estado de amadurecimento que permitem elucidar com grande evidência possíveis candidatos de pequenos RNAs para estudos biológicos. Portanto a descrição destes ácidos nucléicos torna-se importante fonte de conhecimento para futuros estudos

averiguarem a participação destas moléculas na regulação da expressão gênica de fatores de virulência da uropatogênese de *P. mirabilis*.

1.3 EMBASAMENTO TEÓRICO

O embasamento teórico está constituído de 3 etapas:

- a) Revisão da literatura
- b) Estudo e aplicação dos programas de bioinformática para identificação de ncRNAs.
- c) Análise e anotação manual dos ncRNAs

A revisão da literatura propiciou o embasamento para as respostas das questões chaves iniciais: existem ncRNAs codificados por *P. mirabilis*? Se sim, quantos são? Quais são caracterizados?

A análise manual dos resultados possibilitou uma análise rápida e correta dos dados.

2. DESENVOLVIMENTO

2.1. REVISÃO DA LITERATURA

2.1.1 Infecção do trato urinário (ITU)

O trato urinário representa um sítio frequente de infecções tanto em pacientes da comunidade como em pacientes internados em unidades hospitalares, sendo comuns em todas as idades, e estão entre as causas que mais requerem tratamento ambulatorial (MASSON et al., 2009; RORIZ-FILHO et al., 2010).

A infecção do trato urinário caracteriza-se pela multiplicação da bactéria em qualquer parte do trato urinário, seja nos rins, ureteres, bexiga ou uretra. Indivíduos com alto risco de ITU incluem neonatos, mulheres sexualmente ativas e homens e mulheres que estejam utilizando cateter uretral (MOBLEY, BELAS, 1995). Pode ser classificada, de acordo com o local de infecção, como cistite, quando na bexiga, pielonefrite, quando nos rins, e bacteriúria, quando há presença de bactérias na urina. Esta enfermidade pode ser assintomática ou sintomática, tendo ampla variação de sintomas que vão desde leve irritação ao urinar, sepse, podendo até mesmo levar à morte (FOXMAN, 2002; MASSON et al., 2009).

Apesar de todos estarem sujeitos à ITU, existem subpopulações com risco aumentado desta infecção incluindo crianças (LO et al., 2010), gestantes, idosos, diabéticos (MARQUES et al., 2012), pacientes com lesão medular (PANNEK, 2011), e pacientes cateterizados (SEDOR, MULHOLLAND, 1999). Exceto nos primeiros meses de vida, o sexo feminino é mais vulnerável à ITU do que o masculino (HEILBERG, SCHOR, 2003).

A infecção do trato urinário associada a cateter (ITU-CA) é a infecção nosocomial mais comum dentre todas, e representa mais de 80% das ITU nosocomiais (RICHARDS et al., 1999; JACOBSEN et al., 2008). Pacientes cateterizados são vulneráveis à infecção por que o dispositivo reduz a defesa de uma bexiga normal. O enchimento e esvaziamento da bexiga saudável ajuda na limpeza do trato urinário, o que não ocorre com esses pacientes, onde a urina apenas escorre para o saco coletor. Isso permite que ocorra a migração bacteriana da pele contaminada no sítio de inserção. Na bexiga, ocorre um sistema de cultura contínuo, onde as bactérias invadem o reservatório de urina reabastecido pelos rins, o que permite rápida multiplicação gerando cerca de 10^8 UFC/ml (STICKLER, 2014).

A maioria desses patógenos tem origem fecal ou são residentes da pele da própria microbiota nativa dos pacientes. A entrada dessas bactérias pode acontecer no momento da inserção do cateter, através da luz do cateter, ou ao longo da interface entre o dispositivo e a uretra (JACOBSEN et al., 2008).

2.1.2 *Proteus* sp.

O gênero *Proteus* pertence à família Enterobacteriaceae e são amplamente distribuídos na natureza. O traço marcante que o diferencia de outros membros desta família é o “swarming”, uma diferenciação celular que facilita a movimentação e é importante no processo infeccioso (WILLIAMS; SCHWARZHOF, 1978; MOBLEY; BELAS, 1995). Bactérias deste gênero são bacilos Gram negativos, dimórficos, anaeróbio facultativos, móveis com flagelos peritríqueos. Esses microrganismos exibem atividade proteolítica sob condições de aerobiose e anaerobiose facultativa. A desaminação oxidativa de aminoácidos e a habilidade de hidrolisar ureia com a produção de amônia e dióxido de carbono são as propriedades bioquímicas mais representativas deste gênero (NIELUBOWICZ; MOBLEY, 2010).

O gênero *Proteus* é composto por quatro espécies: *P. mirabilis*, *P. vulgaris*, *P. penneri* e *P. myxofaciens* (MANOS, BELAS, 2006), não sendo relacionadas, a última, com infecções em humanos (O'HARA et al., 2000). A espécie mais comum é *P. mirabilis* que, de longe, é o membro mais estudado (MANOS, BELAS, 2006).

P. mirabilis não é um comum causador de ITU em pacientes normais, sendo mais relacionado à ITU complicada, principalmente à cateter (CHEN et al., 2012). Essa bactéria tem, pelo menos, quatro adesinas diferentes que auxiliam na ligação à superfícies, e uma cápsula de exopolissacarídeos que as protege contra as defesas do hospedeiro. São capazes, também, de secretar α -hemolisinas, proteínas de captura de ferro, proteases e aminoácido desaminases, as quais são importantes para aquisição de nutrientes dos tecidos e fluidos. Uma protease de imunoglobulina (Ig) A é produzida e degrada Ig secretadas no muco das superfícies epiteliais. Além disso, produzem uma potente urease e conseguem migrar rapidamente sobre superfícies sólidas (STICKLER, 2014).

2.1.3 Genômica de *Proteus mirabilis*

Numa consulta realizada na data 22/05/16 ao sítio <http://www.ncbi.nlm.nih.gov/genome/genomes/1162>, foram observados 55 genomas de *P. mirabilis* sequenciados, sendo que somente quatro destes possuem a anotação do genoma completo. Deste total de genomas são relatados 09 plasmídios. Neste sítio também observa-se que o tamanho do genoma varia de 3,0377 a 5,0252 Mb.

O primeiro genoma sequenciado foi da cepa uropatogênica *P. mirabilis* HI4320. O seu genoma é de 4,063 Mb de tamanho e há um único plasmídeo que consiste de 36.289 nucleotídeos. Anotação identificou 3.685 sequências de codificação e sete *loci* de rRNA. A análise da sequência confirmou a presença de vários genes de virulência como toxinas, enzimas, flagelo e 17 tipos de fímbrias (PEARSON et al., 2008).

2.1.4 Pequenos RNAs (sRNAs)

Os sRNAs foram descritos inicialmente em procariotos, sendo que o primeiro foi descrito em *E. coli*, é transcrito no cromossomo e possui 174 nucleotídeos (MIZUNO, et al., 1984). Na época não foi relatada nenhuma importância fisiológica do mesmo.

Décadas depois, em 2001 e 2002, foram relatadas a identificação de vários novos sRNAs através de pesquisas computacionais sistemáticas buscando sequências conservadas de promotores órfãos e terminadores nas regiões intergênicas do genoma de *E. coli* (LIVNY ; WALDOR, 2007).

Os sRNAs podem ter funções estruturais, catalíticas ou regulatórias (EDDY, 2001). Um importante exemplo do papel fisiológico destas moléculas é fornecido pela caracterização de alvos, ou seja, sequências ou moléculas com as quais os sRNAs interagem; que indicam a atuação destas moléculas como reguladores pós-transcricionais (WAGNER; ROMBY, 2015).

As moléculas de sRNAs atuam através do seu pareamento com sequências alvo por meio de complementaridade específica ou parcial de bases (GOTTESMAN e STORZ, 2011). Assim, conseguem inibir a transcrição ao bloquear sítios de ligação aos ribossomos, e induzir a transcrição de mRNAs, impedindo a

formação de estruturas secundárias em sua região 5' não traduzida (5' UTR) (STORZ *et al.*, 2011).

Em procariotos os sRNAs desempenham papéis fundamentais em redes reguladoras de expressão nas respostas à estímulos ambientais, inclusive em bactérias patogênicas como *Vibrio cholerae*, *Salmonella sp*, *Staphylococcus aureus*, *Streptococcus agalactiae* e *E. coli* (LENZ *et al.*, 2004; SITTKA, 2008; FELDEN *et al.*, 2011, 238 PATENGE *et al.*, 2015).

2.1.5 Análise *in silico* para localização de sRNAs

A abordagem computacional é considerada uma das mais eficientes para localização de novos sRNAs (EDDY, 2001). Podendo ser dividida de acordo com o método de busca utilizado; dentre os quais, podem ser citados os que procuram estruturas secundárias consensuais; os que efetuam a busca por sinais de transcrição raros; e os que aplicam genômica comparativa e *ab initio*, que independem da busca por similaridade e de estruturas secundárias (SRIDHAR e GUNASEKARAN, 2013).

2.1.5.1 INFERNAL

INFERNAL ("INFERENCE of RNA ALIGNment") é um *software* livre utilizado para pesquisar em bancos de dados de sequências de DNA a estrutura de RNA em sequência semelhantes. Este implementa um caso especial de gramáticas livres de contexto de perfis estocásticos chamados modelos de covariância (CM). Um CM é como um perfil de sequência, mas obtém uma combinação da sequência de consenso e de RNA da estrutura secundária deste consenso, de modo que, em muitos casos, é mais capaz de identificar homólogos de RNA que conservam a sua estrutura secundária do que a sua sequência primária (NAWROCKI; EDDY, 2013).

2.1.5.2 NOCORNAC

NOCORNAC (non-coding RNA characterization) é um programa para a predição e caracterização dos transcritos ncRNA em bactérias (HERBIG; NIESELT, 2011). NOCORNAC utiliza as coordenadas do suposto loci ncRNA como entrada para então realizar a predição. Este programa não se limita à abordagem das

regiões intergênicas, mas também prediz transcritos codificados de RNA em cis. Para a detecção de transcritos no final 3' este software integra o programa TransTermHP (KINGSFORD et al., 2007) para prever os sinais de terminação das regiões Rho-independentes. Para a predição de início 5' em regiões desestabilizadas no DNA genômico, é utilizado o chamado modelo SIDD (Stress Induced Duplex Destabilization) (BENHAM; BI, 2004).

2.1.5.3 Rfam

Rfam é uma base de dados de famílias RNA não codificante (GARDNER, et al., 2009). Cada família é composta por um alinhamento múltiplo de sequências (MSA) de um representativo conjunto de sequências, um modelo de covariância (CM), da sequência e estrutura conservada da família construída a partir do MSA, e de um conjunto de putativos homólogos identificados em uma sequência da base de dados European Nucleotide Archive (ENA) (COCHRANE et al., 2013).

Além da anotação do ncRNA, Rfam atribui um tipo para cada família (como Gene, cis-regulatory elemento, etc.), fornece referências para cada família, bem como outras informações relevantes, tais como o Protein Data Bank (PDB) (BERMAN et al., 2000), miRBase (KOZOMARA e GRIFFITHS-JONES, 2014) e o dados do ENA.

2.2. MATERIAL E MÉTODOS

Toda a metodologia foi realizada *in silico*. Foram utilizados programas livres e banco de dados de genomas. As análises foram realizadas no sistema operacional Ubuntu 14.04. Todo procedimento seguiu a seguinte linha de raciocínio:

- A) Aquisição do genoma bacteriano de *P. mirabilis* no banco de genomas GenBank (NCBI) <http://www.ncbi.nlm.nih.gov/genbank/>.
- B) Inferência de RNAs utilizando-se o programa INFERNAL (versão 1.1, extraída do sítio <http://eddylab.org/inferral/>).

- C) Inferência de pequenos RNAs não codificantes com auxílio do programa nocoRNAC (versão 1.23, adquirido do sitio http://it.inf.uni-tuebingen.de/?page_id=167).
- D) Caracterização dos ncRNAs inferidos através da busca pela sequência no sitio Rfam (<http://rfam.xfam.org/>).
- E) Anotação manual e inspeção dos resultados

2.2.1 Sequência de genoma bacteriano

A sequência do genoma da cepa uropatogênica *P. mirabilis* HI4320 foi selecionada em 08 de fevereiro de 2016, no *National Center for Biotechnology Information* (NCBI, 2013). Este genoma possuía o level “Complete Genome”.

2.2.2 Anotação

Para a anotação de sRNAs foi utilizado o programa *cmsearch* do pacote de programas INFERNAL v1.1 - *INFERENCE of RNA aLignment* - (NAWROCKI; EDDY, 2013) com os modelos de covariância do banco de dados Rfam v12.0 (NAWROCKI 2015).

O programa *non-coding RNA characterization* – nocoRNAC v1.23 (HERBIG; NIESELT, 2011), buscou por possíveis regiões promotoras e terminadoras em torno dos resultados do *software cmsearch*.

Todos os resultados foram combinados e convertidos para o formato de anotação genômica GFF através de *scripts* na linguagem de programação *Python* e *Shell Script*.

2.3. RESULTADOS E DISCUSSÃO

Os candidatos a sRNA foram preditos utilizando-se a estratégia mostrada na Figura 1. O genoma de *P. mirabilis* HI4320 foi extraído do banco de genomas NCBI-Genbank. Este foi então analisado pelo programa INFERNAL v1.1/Rfam v12.0, na

busca pelos sRNAs, sendo que os seguintes parâmetros foram excluídos após a análise final: notrunc, rRNA e No Bit Score. Após esta filtragem sobraram 385 loci.

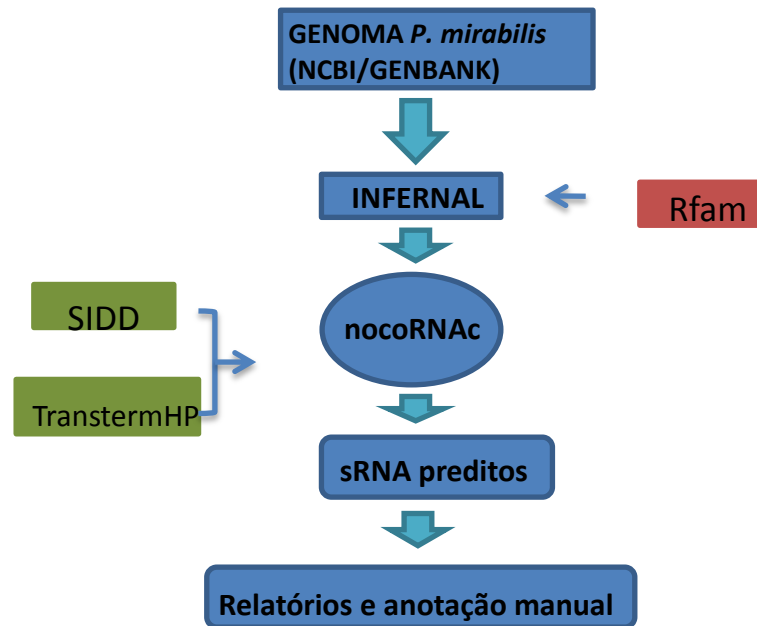


Figura 1- Fluxograma dos procedimentos utilizados para identificar os genes de sRNA no genoma de *P. mirabilis*

Estes possíveis candidatos foram utilizados como resultado de entrada para o programa nocoRNAc. Combinando os dados de SIDD e TransTermHP, este programa gerou a lista final dos candidatos a sRNA. Dos 385 loci analisados o nocoRNAc fez a predição de 39 transcritos de sRNA, sendo que destes 10 tiveram seqüências idênticas dentro deste grupo de 39 transcritos, os quais foram excluídos, totalizando-se então 29 sRNA (Figura 2). O tamanho dos sRNAs variaram de 41 a 505 nucleotídeos, com média de 273 nucleotídeos. Segundo Gottesman e Storz (2011), os sRNAs de procariotos possuem um tamanho aproximado entre 50 e 500 nucleotídeos. Gottesman e Storz (2011), comentam também que devido à rápida evolução nas técnicas de sequenciamento de DNA e de análises *in silico* é difícil estimar um número exato de sRNAs em uma bactéria e que em *E. coli* por exemplo, a variação de 80 a 100 geralmente é citada.

ncRNA_locus254_1
caccaaaataaataaattctacggattgtcaaaatataaagggcagtttccgacattcaatagcattccctattaaaaaacaataagaggaaaattccctca

ncRNA_locus75_1
tattagttgattgtttaataagataatgtttcctgttaaatagaaaaggcactgcaatatgattacagtgccttttctatta

ncRNA_locus209_1
tcaaacaaatcgcagtagctacgtgataacataaattaggttttaaataggtttatcaaaaagatagcctcaccagggaaactatcataataataaata
gaaaaggcactgtaatcatattgcagtgcc

ncRNA_locus247_1
atagttaatcaatccataatagtagctcactctttttgccattctcttttctttactctattcttaagaggcttgtttatttcgaaagaaaagtaagaacgacaat
agcctccaaaaatggaggc

ncRNA_locus249_1
tgcaccttttagctaattagctaattagctaaatcaatagaacacataaaaaaccttcggtcagagcgaagg

ncRNA_locus140_1
gcttaatcaattattgaaaaaataaattttaaagttgtaacaagggtgaataaacttttattccaccttg

ncRNA_locus225_1
tcattataaaaaatggttgaataattatagtaataacaacaatagcttaataaatgaatgaataacaacgctagcagggtgcgaccgtgactacggtcacact
gcta

ncRNA_locus165_1
cctttgatgttctctgttaggcaagctgtaataaagattaaatcgctatttattcactctatctattctataaagaaagggccaatattgcc

ncRNA_locus81_1
ataaattctatgacttaatagcgtcattggttggaaagtacagacctattatgttagtcacggttttagtagttcaccgtttagcgtatccttgatacctctcgttgcgaat
tcctcaaatagtagccataagtaaaaaatccgaaatcaaaccgcaatagccttatggcataactaaattattaaggaattctatgtctaataatgactggta
cagtaaatggttaacgacgataaaggccttggccttaccacctaagatggtagcaaatgattttgtcactttcagcaattcaaaagcgatagctttaaact
actgaaagaaggccaagaatgattctctattgaaaatggtgcaaaaggccagcagctgtaattgtatcggtctgtaataattatggattattgcgtctaaga
cttaaagacactatataagtgatcttactagacgaaataacttaaaaacccactgatgtgagg

ncRNA_locus78_1
ggtttataaaaagcatggattgtgataatttatatacaaaattatcaaca

ncRNA_locus122_1
aagttacacataatgagaatgtctcaatgtaaatatcgcttttccgcttcttttacttttagcgcctaattttgggggcttttgcgctaagaaaagaacgaaaag
tagcgcgggagcctccattttggaggc

ncRNA_locus138_1
taatatttaccctttagagttgaaacggtgtataatcaatacctgtaatatattagatggcaattatctaataataatttaaagatgaagacaacaaaagagaggg
cggttttgaccgccttc

ncRNA_locus28_1
tattttaccctttagagttgaaacggtgtataatcaatacctgtaatatattagatggcaattatctaataataatttaaagatgaagacaacaaaagagagggcgg
ttttgaccgccttc

ncRNA_locus155_1 cttggcgatagaggttttccgacaaaggacacgacattgctcacattgctccagtatttttactagccagctcgggtgctggc

ncRNA_locus41_1
ggataatagcattattttagataataataataaaaattcatctctgcataaatgccaataagagagaggatttaatacactatttctacaaaaaaatgccga
ctggttaaaaagtcgggca

ncRNA_locus166_1
tggcaccgcttgcattatattaatcagttgctcattcaactttttagtgcgtccacaatatgggtgggaacataaccacataagccaagaatgacgccagagtaa
gggtcctaccgtccaacatcatgatactgcctcgggccttcaaacacagggcgacacgtggagtgaggccacccttcttcttttagagagaaaagag
atgaatattgacgagcgggaatatctcagtaagatattccgctctcca

ncRNA_locus207_1 agtaatagcaaaaataaaaagaaggttaattaatataatattaatagacgcccatactaaaaataatttttggatgggcgtcta

ncRNA_locus180_1
atgatcatcctctgtgtctttatgctgtattttggaattagtaattaattttattatcaatgtattcaaaaaaagcagagccaatagctctgc

ncRNA_locus273_1
tatgatcatcctctgtgtctttatgctgtattttggaattagtaattaattttattatcaatgtattcaaaaaaagcagagccaatagctctgc

ncRNA_locus90_1
 ttaataattgttatgctctgcccgggtgcttgatcaatctgggtcaagca

ncRNA_locus173_1
 aataaattgctattctatcgtaagagggcgagtgatgactctctgtataatgaaaaattcctttgactggccgatagcaatatcggc

ncRNA_locus74_1
 cccaataaaggggtgtttgtatatcaaaaccccaataataaattcaactattggggg

ncRNA_locus250_1
 aatgatggatagctttatagaataagtaaggagagatagtgactttataaacataaaaaaacagaggcaataatgcctctg

ncRNA_locus131_1
 aaaagcttgacctactaaaggctaataaaatcaggattttggacaaaaataacctgcataagcaggca

ncRNA_locus5_1
 atagggtattttattagtaaaacagcgctaaatttagcgctgtttttta

ncRNA_locus205_1
 ttgaaaaaaagataaaaaaacagcgctaaatttagcgctg

ncRNA_locus163_1
 atatttagcgacgatgaataatctttctccgctacgtcccctcgcgactaatggaaatattaacttaaccagtttaaacacttcattagccgcggaattgtcgca
 actagagagactaaataaccatcaccacataaaaacataacgactcattttgaaatagcactttctaaaaatagcactttctagtagaaaaggggctacggg
 cccc

ncRNA_locus134_1
 gttgtaattatagaagttatgttttttagatacagataacagacaaacacatcattacattaattgataggtaacccatgatcatcactactctactacaaactctac
 gactaacagcgcatcgcggctgtggtgtgtgctgtggtgtgctgcgcaaaagcgccgtaacgggtccgatcaacacagtttcgcaaccccgcgg
 cgtattaaccgggcggg

ncRNA_locus162_1
 ttgctattttgtgctgtatgatggaatagcttcatcctattcgctatgtaatgataatcaaaaaagcgagtagaaaagttccctggtgtggcgcagtattcg
 cgcacccaacctcggttggg

Figura 2- Genes dos pequenos RNAs não codificadores preditos para *P. mirabilis* uropatogênico

Muitos estudos predizem e analisam sRNAs somente em regiões intergênicas (LIVNY *et al.*, 2005; SRIDHAR *et al.*, 2010; HERSHBERG *et al.*, 2003). No entanto, vários estudos têm demonstrado que os sRNAs parcialmente codificados nas fitas não codificantes de fases de leitura abertas (ORFs) desempenham um importante papel na regulação de vias metabólicas no nível pós-transcricional (KAWANO *et al.*, 2005; De LAY e GOTTESMAN, 2010). Similares resultados foram observados no presente estudo dos 29 sRNA, 18 foram antisense com mRNA e 11 candidatos foram encontrados em regiões intergênicas.

As 29 sequências de DNA de sRNA foram transcritas *in silico* através do sítio http://in-silico.net/tools/biology/sequence_conversion. Então estas foram usadas numa busca no site Rfam (<http://rfam.xfam.org/>), realizada no dia 20/07/16. Segundo a tabela 01 dos 29 sRNA somente oito apresentaram sequência já descrita na literatura. Quanto às famílias quatro são cis-regulatórias e quatro são genes-sRNA e todos se localizam no cromossomo bacteriano.

As sequências ncRNA_locus247_1 e ncRNA_locus122_1 correspondem ao ncRNA Phe-leader. Esta sequência foi identificada em *Lactococcus lactis* através de estudos de bioinformática. Sua função é suprimir a expressão de genes ricos em códons para fenilalanina. Este RNA funciona como peptídeo líder contendo na sua sequência muitos códons para fenilalanina. Normalmente, este ncRNA atua suprimindo a expressão dos genes à jusante a ele (WEINBERG et al., 2010; KOK., 1996).

As demais sequências a seguir, foram primeiramente descritas em *E. coli*.

O locus ncRNA_locus81_1 corresponde ao ncRNA cspA mRNA 5' UTR. Esta sequência é descrita como uma região não traduzida do gene *cspA*. Este ncRNA age como um “RNA termômetro”, regulando a expressão do gene *cspA* em resposta à variação de temperatura (GIULIODORI et al., 2010). Este gene é importante na regulação “cold shock” em Enterobactérias, ao regular a temperatura, a proteína CspA promove a homeostasia da célula bacteriana (YAMANAKA et al., 1999).

Já o locus ncRNA_locus155_1 corresponde ao ncRNA RyhB que tem como função regular o estoque de ferro e o uso deste por proteínas. Este gene ncRNA foi identificado em dois estudos independentes, sendo chamado RyhB por Wassarman et al. (2001) e SraJ por Argaman et al (2001). Estes trabalhos verificaram que os níveis de RyhB-RNA são inversamente correlacionados com os níveis de mRNA para o operon *sdhCDAB* (succinato desidrogenase). Outros genes têm sido previstos computacionalmente e verificado como alvos deste ncRNA através de análise por microarranjo (TJADEN et al., 2006)

O locus cRNA_locus166_1 é o sRNA GlmZ também conhecido como SraJ. Este ncRNA foi descoberto em *E. coli* através de estudos *in silico* e confirmado por *Norther blotting* (ARGAMAN et al., 2001). Sua função é regular positivamente o mRNA Glms (KALAMORZ et al., 2007; URBAN et al., 2007).

A sequência ncRNA_locus173_1 corresponde ao RNA sroB, o qual é encontrado em várias enterobactérias (VOGEL et al., 2003). Este regula negativamente a proteína YbfM, ligando-se ao mRNA *ybfM* por interação antisense (RASMUSSEN et al., 2009).

Já ncRNA_locus162_1 corresponde ao ArcZ que é um regulador antisense de vários genes de proteínas (ARGAMAN et al., 2001). Papenfort et al 2009, demonstraram que em *Salmonella* este RNA regula a expressão dos genes sdaCB (envolvidos na captação de serina) e tpx (envolvido no estresse oxidativo), ligando-se ao ribossomo.

E por último, ncRNA_locus163_1 foi caracterizado como SraG cuja função ainda não é conhecida (ARGAMAN et al., 2001).

Tabela 01- Caracterização das sequências de ncRNA pelo nome, tipo, número e regulação na célula bacteriana.

Locus nome descrito Rfam	Tipo de RNA	Número Rfam	Regulação	Referência
ncRNA_locus247_1 Phe-leader	Cis-reg	RF01743	Suprime a expressão de genes ricos em códons para fenilalanina	WEINBERG et al., 2010
ncRNA_locus81_1 cspA mRNA 5' UTR	Cis-reg; termoregulador	RF01766	Regula a expressão do gene <i>cspA</i> em resposta à variação de temperatura	GIULIODORI et al., 2010
ncRNA_locus122_1 Phe-leader	Cis-reg	RF01743	Suprime a expressão de genes ricos em códons para fenilalanina	WEINBERG et al., 2010
ncRNA_locus155_1 RyhB	Cis-reg	RF00057	Regula o estoque de ferro e o uso deste por proteínas	WASSARMAN et al. 2001

ncRNA_locus166_1 GlmZ	Gene; sRNA	RF00083	Ativador do gene <i>glmZ</i>	ARGAMAN et al., 2001
ncRNA_locus173_1 sroB	Gene; sRNA	RF00368	Regula negativamente a proteína YbfM	VOGEL et al., 2003
ncRNA_locus163_1 SraG	Gene; sRNA	RF00082	Função desconhecida	ARGAMAN et al., 2001
ncRNA_locus162_1 ArcZ	Gene; sRNA	RF00081	Regulador antisense de vários genes de proteínas	ARGAMAN et al., 2001

3. CONCLUSÃO

Através das análises *in silico* foi possível prever 29 pequenos RNAs não codificadores no genoma de *P. mirabilis* uropatogênico. Do total oito possuem função já conhecida. Estes resultados são inéditos e podem contribuir no conhecimento da regulação da expressão gênica deste uropatógeno. Para que estas moléculas deixem de ser somente previstas uma futura validação experimental *in vitro* torna-se necessária bem como a busca pelos seus salvos e o entendimento de suas funções regulatórias.

REFERÊNCIAS

- ARGAMAN, L et al. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. **Curr Biol**. v.11, n.12, p. 941–950. 2001.
- BENHAM, C.J, BI, C. The analysis of stress-induced duplex destabilization in long genomic DNA sequences. **J Comput Biol**, v.11,n.4,p.519-543, 2004.
- BERMAN, H.M. et al. The Protein Data Bank. **Nucleic Acids Res**. v.28, p.235–242, 2000.
- CHEN, C. Y., et al. *Proteus mirabilis* urinary tract infection and bacteremia: risk factors, clinical presentation, and outcomes. **J. of Microb.Immun Inf.**, v. 45, n. 3, p.228-236, 2012.
- COCHRANE,G., et al. Facing growth in the European nucleotide archive. **Nucleic Acids Res.**,v.41, p.30–35, 2013.
- EDDY, S. R. Non-coding RNA genes and the modern RNA world. **Nature Reviews Genetics**, v. 2, p.919-929, 2001.
- FELDEN B, et al. The *Staphylococcus aureus* and its commitment to virulence. **PLoS Pathog**. v.7, p.1–10, 2011.
- FOXMAN, B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. **The Am. J. Medicine**, v. 113, n. 1, p. 5–13, 2002.
- GARDNER, P.P. et al. Rfam: updates to the RNA families database. **Nucleic Acids Res**. v.37, p.136-140, 2008.
- GIULIODORI, A.M. et al. The *cspA* mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. **Mol. Cell**. v. 37, n1, p. 21–33, 2010.
- GOTTESMAN, S.; STORZ, G. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. **Cold Spring Harbor perspectives in biology**, v. 3, p.305 n.12, 2011.
- HEILBERG, I. P.; SCHOR, N. Abordagem diagnóstica e terapêutica na infecção do trato urinário – Itu. **Revista da Associação Médica Brasileira**, São Paulo, v. 49, n. 1, p. 109-116, 2003.
- HERBIG, A, NIESELT, K. nocoRNAC: characterization of non-coding RNAs in prokaryotes. **BMC Bioinformatics**. v.12, p.40, 2011.
- JACOBSEN, S. M. et al. Complicated catheter-associated urinary tract infections due to *Escherichia coli* and *Proteus mirabilis*. **Clin. Microb. Reviews**, v. 21, n.1, p. 26–59, 2008.

KALAMORZ, F. et al. Feedback control of glucosamine-6-phosphate synthase GImS expression depends on the small RNA GImZ and involves the novel protein YhbJ in *Escherichia coli*. **Mol Microbiol.** v.65, n.6, p.1518–1533, 2007.

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

KOK, J. Inducible gene expression and environmentally regulated genes in lactic acid bacteria. **Antonie Van Leeuwenhoek.** v.70, n.3 p.29–45, 1996.

KOZOMARA, A., GRIFFITHS-JONES, S. miRBase: annotating high confidence microRNAs using deep sequencing data. **Nucleic Acids Res.**, v.42, p.68–73, 2014

LENZ, D.H., et al. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. **Cell.**;v.118, n 1093, p.69–82, 2004.

LIVNY, J., WALDOR, M.K. Identification of small RNAs in diverse bacterial species. **Curr. Opin. Microbiol.** v.10, p.96–101, 2007.

LO, D. S.; et al. Infecção urinária em menores de 15 anos: etiologia e perfil de sensibilidade antimicrobiana em hospital geral de pediatria. **Revista Paulista de Pediatria**, São Paulo, v. 28, n. 4, p. 299-303,2010.

MANOS, J.; BELAS, R. The genera *Proteus*, *Providencia*, and *Morganella*. **Prokaryotes**, v. 6, p. 245–269, 2006.

MARQUES, L. P. J. et al. Epidemiological and clinical aspects of urinary tract infection in community-dwelling elderly women. **The Brazilian J of Infec Dis**, v. 16, n. 5, p. 436-41, 2012.

MASSON, P.; MATHESON, S.; WEBSTER, A. C.; CRAIG, J. C. Meta-analyses in prevention and treatment of urinary tract infections. **Infectious Disease Clinics of North America**, v. 23, n. 2, p. 355–385, 2009.

MIZUNO, T., CHOU, M.Y., INOUE, M. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (miRNA). **Proc. Natl. Acad. Sci. USA**, v.81, p.1966–1970, 1984

MOBLEY, H. L. T.; BELAS, R. Swarming and pathogenicity of *Proteus mirabilis* in the urinary tract. **Trends in Microbiology**, v. 3, n. 7, p. 280-284, 1995.

NAWROCKI, E.P, EDDY, S.R. Infernal 1.1: 100-fold faster RNA homology searches. **Bioinformatics**, v.29, p 2933–2935, 2013

NCBI. Database resources of the National Center for Biotechnology Information. **Nucleic Acids Res.**v.41, p.8–20, 2013.

NIELUBOWICZ, G.R, MOBLEY, H.L. Host-pathogen interactions in urinary tract infection. **Nat. Rev. Urol.** v.7, p.430–441, 2010.

O'HARA, C. M. et al Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. **Clin. Microb. Reviews**, v. 13, n. 4, p. 534–546, 2000.

PANNEK, J. Treatment of urinary tract infection in persons with spinal cord injury: guidelines, evidence, and clinical practice - A questionnaire-based survey and review of the literature. **The Journal of Spinal Cord Medicine**, v. 34, n. 1, p. 11–15, 2011.

PAPENFORT, K. et al. Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. **Mol Microbiol.** v.74, n.1, p.139–158, 2009.

PATENGE, N, et al. Genome-wide analyses of small non-coding RNAs in streptococci. **Front Genet**; v.6, p.1–13, 2015.

PEARSON, M.M. et al. Complete genome sequence of uropathogenic *Proteus mirabilis*, a master of both adherence and motility. **J. Bacteriol**, v.190, n.11, p.4027-4037, 2008.

RASMUSSEN, A.A. et al. A conserved small RNA promotes silencing of the outer membrane protein YbfM. **Mol. Microbiol.** v.72, n.3, p.566–577, 2009.

RICHARDS, M. J. et al. Nosocomial infections in medical intensive care units in the United States. National Nosocomial Infections Surveillance System. **Critical Care Medicine**, v. 27, n. 5, p. 887-92, 1999.

RORIZ-FILHO, J. S et al.. Infecção do trato urinário. **Medicina (Ribeirão Preto)**, v. 43, n. 2, p. 118-25, 2010.

SEDDOR, J.; MULHOLLAND, S. G. Hospital-acquired urinary tract infections associated with the indwelling catheter. **Urol. Clin of North Am.**, v. 26, n.4, p. 821-8, 1999.

SITTKA, A, et al. Deep Sequencing Analysis of Small Noncoding RNA and mRNA Targets of the Global Post-Transcriptional Regulator, Hfq. **PLoS Genet.** v.4, n.8, p.01-20, 2008.

SRIDHAR ,J., GUNASEKARAN, P. Computational small RNA prediction in bacteria. **Bioinform Biol Insights.** v.7, p.83–95. 2013

STICKLER, D. J. Clinical complications of urinary catheters caused by crystalline biofilms: something needs to be done. **Journal of Internal Medicine**, v. 276, n. 2, p.120-9, 2014.

STORZ, G.; VOGEL, J.; WASSARMAN, K. M. Regulation by Small RNAs in Bacteria: Expanding Frontiers. **Molecular Cell**, v. 43, n. 6, p. 880–891, 2011.

TJADEN, B. et al. Target prediction for small, noncoding RNAs in bacteria. **Nucleic Acids Res.** v.34, n.9, p.2791–2802, 2006.

URBAN, J.H. et al. A conserved small RNA promotes discoordinate expression of the glmUS operon mRNA to activate GlnS synthesis. **J Mol Biol.** v. 373, n.3, p. 521–528, 2007.

VOGEL, J. et al. RNomics in Escherichia coli detects new sRNA species and indicates parallel transcriptional output in bacteria. **Nucleic Acids Res.** v.31, n.22, p.6435–6443, 2003.

WAGNER, E. G. H.; ROMBY, P. Small RNAs in Bacteria and Archaea. **Advances in Genetics.** v. 90. p.133–208, 2015.

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

WEINBERG, Z. et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea and their metagenomes. **Genome Biol.** v.11, n.3, p.01-17, 2010.

WILLIAMS, F. D.; SCHWARZHOF, R. H. Nature of the swarming phenomenon in Proteus. **Ann. Rev. of Microb.**, v. 32, p. 101–122, 1978.

YAMANAKA K, MITTA M, INOUE M. Mutation analysis of the 5' untranslated region of the cold shock cspA mRNA of Escherichia coli. **J. Bacteriol.** v.181, n.20, p.6284–6291, 1999.