



**UNIVERSIDADE FEDERAL DA BAHIA**  
**INSTITUTO DE CIÊNCIAS DA SAÚDE**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA**

**DANILO JOBIM PASSOS GIL DA ROCHA**

**DESENHO E VALIDAÇÃO DE UM CONJUNTO DE PRIMERS UNIVERSAIS**  
**BACTERIANOS PARA NORMALIZAÇÃO DE ENSAIOS DE PCR**  
**QUANTITATIVA EM TEMPO REAL**

Salvador

2017

**DANILO JOBIM PASSOS GIL DA ROCHA**

**DESENHO E VALIDAÇÃO DE UM CONJUNTO DE PRIMERS UNIVERSAIS  
BACTERIANOS PARA NORMALIZAÇÃO DE ENSAIOS DE PCR  
QUANTITATIVA EM TEMPO REAL**

Trabalho de dissertação apresentado ao Programa de Pós-Graduação em Biotecnologia do Instituto de Ciências da Saúde da Universidade Federal da Bahia como requisito para obtenção do título de Mestre em Biotecnologia.

Orientador: Prof. Dr. Luis Gustavo Carvalho Pacheco

Salvador

2017

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas (SIBI/UFBA),  
com os dados fornecidos pelo(a) autor(a)

Rocha, Danilo Jobim Passos Gil da  
Desenho e validação de um conjunto de primers universais bacterianos para normalização de ensaios de PCR quantitativa em tempo real / Danilo Jobim Passos Gil da Rocha. -- Salvador, 2017.  
76 f.

Orientador: Luis Gustavo Carvalho Pacheco.  
Dissertação (Mestrado - Biotecnologia) -- Universidade Federal da Bahia, Instituto de Ciências da Saúde, 2017.

1. RT-qPCR. 2. Real time PCR. 3. Genes de referência. 4. Bactérias. 5. Biologia Molecular. I. Pacheco, Luis Gustavo Carvalho. II. Título.

# DANILO JOBIM PASSOS GIL DA ROCHA

## Desenho e validação de um conjunto de *primers* universais bacterianos para normalização de ensaios de PCR quantitativa em tempo real

Dissertação apresentada como requisito para obtenção do grau de Mestre em Biotecnologia pelo Instituto de Ciências da Saúde da Universidade Federal da Bahia.

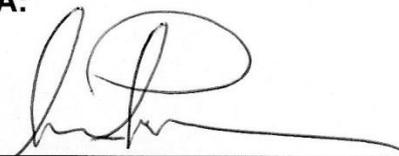
Aprovada em 01 de setembro de 2017.

### BANCA EXAMINADORA:

Luís Gustavo Carvalho Pacheco -Orientador

Doutor em Bioquímica e Imunologia pela Universidade Federal de Minas Gerais, UFMG, Brasil.

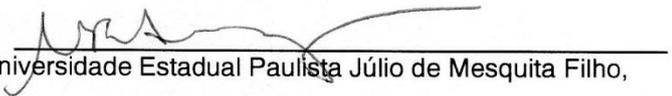
Universidade Federal da Bahia.



Milton Ricardo de Abreu Roque

Doutor em Ciências Biológicas pela Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Brasil.

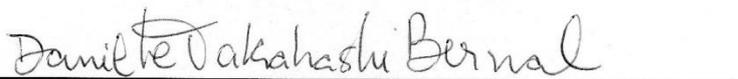
Universidade Federal da Bahia



Daniele Takahashi Bernal

Doutora em Agronomia pela Universidade de São Paulo, USP, Brasil.

Universidade Federal da Bahia.



Dedico esta dissertação aos meus pais, pela vida que me deram, pelo amor e pelo esforço para me fornecer uma educação digna.

## **AGRADECIMENTOS**

Agradeço aos meus pais por me darem a vida, por acreditarem em mim, por me fazer querer ser sempre um ser humano melhor.

A minha companheira de laboratório e amiga Carolina Santos, por sermos de ambos os ombros nas horas difíceis e a alegria nos momentos de leveza.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelos anos de investimento na forma de bolsa de estudos.

Ao Ciências sem Fronteiras, pela oportunidade de estudar no exterior numa das melhores universidades do mundo.

A tantos amigos que ao me perceber em momentos de fadiga tanto física quanto mental, estiveram presentes para me animar e me fazer voltar os trilhos.

Ao meu orientador Prof. Dr. Luis Pacheco, pela oportunidade e pela paciência.

“Amo aqueles que não procuram primeiro  
por motivos atrás das estrelas para sua  
queda e seu sacrifício: mas os que se  
sacrificam pela terra, para que, um dia, a  
terra venha a ser do Super-homem.”

Assim falou Zaratustra,  
Friederich Nietzsche

ROCHA, D. J. P. G. Desenho e validação de um conjunto de primers universais bacterianos para normalização de ensaios de PCR quantitativa em tempo real. 77 f. Il. Dissertação (Mestrado) - Salvador. Instituto de Ciências da Saúde, Universidade Federal da Bahia, 2017.

## RESUMO

O método mais comum de normalização em ensaios de quantificação relativa da expressão gênica por PCR quantitativa em tempo real (qPCR) faz uso de um gene de referência, que deve manter sua expressão estável sob diferentes condições experimentais. Em uma meta-análise da literatura e banco de dados de experimentos de análise de expressão gênica em bactérias, os genes *gyrA* (DNA gyrase subunidade A), *gyrB* (DNA gyrase subunidade B), *dnaG* (DNA primase), *era* (GTPase era) e *secA* (translocase proteica subunidade secA), figuram entre os mais estáveis em diversas condições experimentais. Neste cenário, este estudo se propõe a desenvolver e construir um conjunto de *primers* universais para esses genes de referência a partir de organismos modelo de grupos bacterianos de interesse clínico e biotecnológico. As ferramentas de alinhamento, *ClustalOmega* e PCR virtual, *Primer-Blast* foram utilizadas para encontrar regiões conservadas entre os genes candidatos em diferentes organismos bacterianos. Dentro do complexo *Mycobacterium tuberculosis*, todos os genes apresentaram homologia suficiente para o desenho de *primers* universais, enquanto que em outros grupos bacterianos a homologia de sequência foi restrita a algumas espécies. Potenciais *primers* universais foram desenhados para diferentes grupos bacterianos e os *primers* para a família *Enterobacteriaceae* foram validados por RT-qPCR em *Escherichia coli*; os resultados foram comparados contra o gene comumente usado, *16S rRNA*. Os *primers* testados apresentaram eficiências de amplificação dentro dos limites esperados e as expressões dos genes de referência foram estáveis nas condições estudadas, tendo sido o gene *dnaG* o mais estável, de acordo com os softwares *NormFinder* e *RefFinder*. Conclui-se que é possível desenhar *primers* universais funcionais para normalização de RT-qPCR em grupos bacterianos específicos, contudo o baixo nível de conservação gênica de determinados genes pode limitar suas utilizações.

Palavras-chave: RT-qPCR, real time PCR, genes de referência, *housekeeping genes*, genes de governança, bactérias, normalização.

ROCHA, D. J. P. G. Design and validation of a set of bacterial universal primers for normalization of quantitative real-time PCR assays. 77 f. II. Dissertation (Master degree) - Salvador. Instituto de Ciências da Saúde, Universidade Federal da Bahia, 2017.

### **ABSTRACT**

The most common method of normalization in quantitative real-time quantitative PCR (qPCR) relative quantification of gene expression assays makes use of a reference gene, which should maintain its expression stable under different experimental conditions. In a meta-analysis of literature and bacterial gene expression data banks, the genes *gyrA* (DNA gyrase subunit A), *gyrB* (DNA gyrase subunit B), *dnaG* (DNA primase), *era* (GTPase *era*) e *secA* (protein translocase subunit *secA*), figured among the most stable under many different experimental conditions. In this scenario, this study proposes to develop and build a set of universal primers for these reference genes from model organisms of bacterial groups of clinical and biotechnological interest. The alignment tools, ClustalOmega and virtual PCR, Primer-Blast were used to find conserved regions between candidate genes in different bacterial organisms. Within the *Mycobacterium tuberculosis* complex, all genes showed sufficient homology for the design of universal primers, while in other bacterial groups sequence homology was restricted to some species. Potential universal primers were designed for different bacterial groups and primers for the Enterobacteriaceae family were validated by RT-qPCR in *Escherichia coli*; the results were compared against the commonly used 16S rRNA gene. The primers tested showed amplification efficiencies within the expected limits and the reference gene expressions were stable under the conditions studied, with the *dnaG* gene being the most stable, according to the NormFinder and RefFinder software. It is concluded that it is possible to design functional universal primers for normalization of RT-qPCR in specific bacterial groups; however the low level of gene conservation of certain genes may limit their uses.

Key-words: RT-qPCR, real time PCR, reference genes, *housekeeping genes*, bacteria, normalization.

## LISTA DE TABELAS

Tabela 1 Parâmetros de ciclagem de PCR quantitativa como sugerido pelo fabricante utilizados na reação pro termociclador. ....	35
Tabela 2 <i>Primers</i> de enterobactérias desenhados a partir dos genes <i>dnaG</i> , <i>gyrA</i> , <i>gyrB</i> e <i>secA</i> de <i>Escherichia coli</i> e suas propriedades técnicas.....	44
Tabela 3 Amplificação baseada em PCR virtual com a ferramenta <i>Primer-BLAST</i> de enterobactérias de interesse clínico.....	45
Tabela 4 Amplificação de espécies do complexo <i>Mycobacterium tuberculosis</i> baseada em PCR virtual com a ferramenta <i>Primer-BLAST</i> ..	47
Tabela 5 <i>Primers</i> para complexo <i>Mycobacterium tuberculosis</i> (MTC) desenhados a partir dos genes <i>dnaG</i> , <i>gyrA</i> , <i>gyrB</i> e <i>secA</i> de <i>Mycobacterium tuberculosis</i> e suas propriedades técnicas..	48
Tabela 6 <i>Primers</i> dos genes <i>dnaG</i> , <i>secA</i> , <i>gyrB</i> e <i>16s</i> para diferentes organismos modelo em firmicutes e proteobacteria.....	49
Tabela 7 Dados de quantificação e pureza do RNA obtido. ....	51
Tabela 8 Amostras e os valores de Ct obtidos por qPCR para cada diluição de cada par de <i>primer</i> . ....	54
Tabela 9 Eficiência dos <i>primers</i> para os genes candidatos a referência e a inclinação das retas obtidas a partir dos valores de Ct. ....	54
Tabela 10 Ranqueamento dos melhores genes de referência para os quatro principais métodos de acordo com a ferramenta <i>Ref-Finder</i> .....	57

## LISTA DE FIGURAS

Figura 1 Fluxograma da metodologia de trabalho. ....	26
Figura 2 Página de resultados do alinhamento pela ferramenta BLASTn do NCBI.....	28
Figura 3 Visualização parcial do alinhamento de sequências para o gene <i>gyrB</i> via <i>jalview</i> . ....	39
Figura 4 Visualização de uma região conservada do alinhamento de sequências para o gene <i>gyrB</i> . ....	40
Figura 5 Visualização parcial do alinhamento de sequências para o gene <i>gyrA</i> via <i>jalview</i> . ....	40
Figura 6 Visualização de uma região conservada do alinhamento de sequências para o gene <i>gyrA</i> . A cor azul escuro representa a porcentagem de semelhança.....	40
Figura 7 Visualização parcial do alinhamento de sequência para o gene <i>dnaG</i> via <i>jalview</i> . ....	41
Figura 8 Representação de organismos que partilham maior semelhança para o gene <i>dnaG</i> , segundo <i>clustalOmega</i> , via <i>jalview</i> .....	41
Figura 9 Visualização parcial do alinhamento de sequências para o gene <i>era</i> via <i>jalview</i> .. ....	42
Figura 10 Representação de organismos que partilham maior semelhança para o gene <i>era</i> , segundo <i>clutalomega</i> , via <i>jalview</i> .....	42
Figura 11 Visualização parcial do alinhamento de sequências para o gene <i>secA</i> via <i>jalview</i> . ....	43

Figura 12 Representação de organismos que partilham maior semelhança para o gene <i>era</i> , segundo <i>clutalomega</i> , via <i>jalview</i> .....	43
Figura 13 Visualização parcial do alinhamento do gene <i>gyrB</i> dentro da família <i>Mycobacteriaceae</i> via <i>jalview</i> . .....	46
Figura 14 Visualização de uma região conservada do alinhamento de sequências para o gene <i>gyrB</i> na família <i>Mycobacteriaceae</i> . .....	47
Figura 15 Boxplot dos dados de <i>gyrB</i> , <i>gyrA</i> , <i>dnaG</i> e <i>secA</i> para <i>Escherichia coli</i> .....	50
Figura 16 Boxplot dos dados de <i>gyrB</i> , <i>gyrA</i> , <i>dnaG</i> e <i>secA</i> para <i>Bacillus subtilis</i> .. .....	50
Figura 17 Curvas de diluição obtidas por qPCR para amostras diluídas 1x, 10x, 100x, 1000x e 10000x. ....	53
Figura 18 Curvas de dissociação para os genes <i>gyrB</i> , <i>16S</i> , <i>dnaG</i> e <i>secA</i> .....	55
Figura 19 Boxplot dos dados de Cq entre as amostras das condições normal e as obtidas após estresse térmico de 47°C por 10min. ....	56
Figura 20 Estabilidade de expressão para os genes candidatos a gene normalizador utilizando o software <i>normfinder</i> .....	57
Figura 21 Relação de estabilidade dos genes candidatos a referência entregue pela ferramenta online <i>Ref-Finder</i> .. .....	58

## LISTA DE SÍMBOLOS E ABREVIACIONES

$\Delta$	Delta
$\mu$	Micro
BLAST	<i>Basic Local Alignment Search Tool</i>
cDNA	Ácido desoxirribonucleico complementar
Cq	Ciclo de quantificação
DNA	Ácido desoxirribonucleico
dNTPs	Desoxinucleotídeos trifosfatados
DO	Densidade ótica
EDTA	Ácido etileno diamino tetracético
g	Gravidade
LB	Luria-Bertani
min	Minutos
MIQE	Informações Mínimas para Publicações de Experimentos Envolvendo qPCR em Tempo Real
mL	Mililitro
mRNA	Ácido ribonucleico mensageiro

<b>MTC</b>	<b><i>Micobacterium tuberculosis complex</i></b>
<b>NCBI</b>	<b>National Center for Biotechnology Information</b>
<b>°C</b>	<b>Graus Celsius</b>
<b>PATRIC</b>	<b><i>Pathosystems Resource Integration Center</i></b>
<b>PCR</b>	<b>Reação em Cadeia da Polimerase</b>
<b>pH</b>	<b>Potencial de hidrogênio</b>
<b>qPCR</b>	<b>Reação em Cadeia da Polimerase Quantitativa em Tempo Real</b>
<b>RNA</b>	<b>Ácido ribonucleico</b>
<b>RNase</b>	<b>Ribonuclease</b>
<b>RNA-seq</b>	<b>Sequenciamento de RNA</b>
<b>RT</b>	<b>Transcriptase Reversa</b>
<b>RT-qPCR</b>	<b>Transcrição Reversa seguida de Reação em Cadeia da Polimerase</b>

## SUMÁRIO

1. FUNDAMENTAÇÃO TEÓRICA .....	16
1.1 Transcrição Reversa seguida de PCR quantitativa em tempo real (RT-qPCR).....	16
1.2 Funcionamento da qPCR .....	17
1.3 Tipos de Quantificação .....	18
1.4 Informações Mínimas para Publicação de Experimentos de PCR quantitativo em Tempo Real ( <i>The Minimum Information for Publication of Quantitative Real-Time PCR Experiments</i> ) .....	19
1.5 Genes de Referência .....	20
1.6 Normalização .....	21
2. JUSTIFICATIVA .....	23
3. OBJETIVOS .....	25
3.1 Objetivos Específicos.....	25
4. METODOLOGIA.....	26
4.1 Desenho de <i>Primers</i> .....	27
4.2 Análise <i>in silico</i> da estabilidade dos genes candidatos a genes de referência.....	29
4.3 <i>Primers</i> de para organismos modelo em <i>Proteobacteria</i> e <i>Firmicutes</i> 30	
4.4 Cultivo Bacteriano .....	31
4.5 Extração de RNA .....	32
4.5.1 RNA <i>Cleanup</i> .....	34
4.6 Transcriptase Reversa seguida de PCR Quantitativa (RT-qPCR) .....	34
4.6.1 Transcrição Reversa.....	34
4.6.2 qPCR.....	35
4.6.3 Determinação da eficiência dos <i>primers</i> .....	36
4.6.4 Verificação da quantidade de DNA genômico.....	36
4.6.5 Avaliação da estabilidade dos genes candidatos a referência .....	37
5. RESULTADOS .....	38
5.1 Desenho de <i>Primers</i> .....	38
5.1.1 <i>Primers</i> de Proteobactérias .....	38
5.1.2 <i>Primers</i> para Complexo <i>Mycobacterium tuberculosis</i> (MTC) .....	46
5.1.3 <i>Primers</i> de organismos modelo em <i>Firmicutes</i> e <i>Proteobactéria</i> .	49
5.1.4 Análise <i>in silico</i> da estabilidade dos genes candidatos .....	50
5.1.5 Padronização da extração de RNA.....	51

5.2	RT-qPCR .....	52
5.2.1	Determinação da Eficiência dos <i>Primers</i> .....	52
5.2.2	Especificidade das reações de qPCR através de curvas de dissociação.....	55
5.2.3	Análise de estabilidade dos diferentes genes .....	56
6.	DISCUSSÃO .....	59
6.1	Desenho de <i>Primers</i> .....	59
6.1.1	<i>Primers</i> Universais.....	59
6.1.2	<i>Primers</i> para <i>Enterobacteria</i> .....	59
6.1.3	Inciadores para complexo <i>Mycobacterium tuberculosis</i> (MTC)....	61
6.1.4	<i>Primers</i> de organismos modelo para <i>Firmicutes</i> e <i>Proteobacteria</i>	62
6.2	Análise <i>in silico</i> da estabilidade dos genes candidatos .....	63
6.3	Estabilidade dos genes selecionados em resposta ao estresse térmico em <i>Escherichia coli</i> .....	64
6.4	Extração de RNA e verificação de DNA genômico .....	64
6.5	RT-qPCR .....	65
6.5.1	Eficiência dos <i>primers</i> .....	65
6.5.2	Estabilidade dos genes candidatos a referência .....	66
7.	CONSIDERAÇÕES FINAIS .....	69
8.	REFERÊNCIAS .....	70

## 1. FUNDAMENTAÇÃO TEÓRICA

### 1.1 Transcrição Reversa seguida de PCR quantitativa em tempo real (RT-qPCR)

A principal diferença da qPCR (Reação em Cadeia da Polimerase Quantitativa) para a PCR convencional, desenvolvida por Kerry Mullis em 1983 (MULLIS et al., 1986), dá-se pelo uso de um reagente repórter para combinar os passos de amplificação e a detecção da reação de PCR em um único tubo. Quando cada amostra se encontra na fase exponencial da PCR, os reagentes estão em excesso, o produto está dobrando a cada ciclo, e a quantidade de produto possui correlação linear com a quantidade de material inicial. A técnica é comumente empregada para mensuração de níveis de mRNA em material biológico, utilizando de uma etapa de transcrição reversa e síntese de cDNA (DEAN; UDAWELA; SCARR, 2016; NESTOROV et al., 2013a).

A metodologia tornou-se uma técnica de preferência quando empregada em estudos de expressão gênica de diversos organismos. Sua popularidade cresceu exponencialmente nos últimos anos com dezenas de milhares de artigos em diversos campos de estudo, como agricultura, meio-ambiente, pesquisa médica, todos fazendo referência à utilização de RT (transcrição reversa) –qPCR (TAYLOR et al., 2010).

As respostas celulares para sobrevivência, diferenciação, crescimento, dentre outros aspectos, compõem diferentes padrões de expressão de seus genes. A habilidade de quantificar a expressão destes genes em diferentes condições é o pilar de qualquer estudo de função gênica (BUSTIN, 2000). Desde o surgimento da medicina molecular houve um acréscimo no uso de técnicas capazes de quantificar os níveis de RNA. Alguns métodos estão entre os mais usados em quantificação de níveis transcricionais: *Northern Blotting* e hibridização *in situ*, arranjo de cDNA e a RT-qPCR. Entretanto, apesar da recente adoção de RNA-seq como uma poderosa tecnologia para avaliar dados de expressão em diferentes tratamentos, RT-qPCR é o método mais utilizado, servindo como validação para dados obtidos em outros métodos como observações fenotípicas e dados de expressão protéica em proteômica e *DNA Microarray*. Entre suas vantagens estão: maior sensibilidade, precisão na medição de amostras e rapidez, ao eliminar etapas de manipulação de amostra

como a visualização em gel de agarose. (GACHON; MINGAM; CHARRIER, 2004; JOHNSON et al., 2014; NESTOROV et al., 2013; TAYLOR et al., 2010).

A RT-qPCR é, portanto, uma metodologia *in vitro* para amplificar através de enzimas, sequências determinadas de RNA, permitindo a análise de diferentes amostras como o conteúdo de uma única célula (VANGUILDER; VRANA; FREEMAN, 2008). Pode ainda ser aplicada para discriminar relações entre mRNA, ao caracterizar padrões de expressão, ou ainda analisar sua estrutura. Conforme a tecnologia foi desenvolvida, melhoramentos foram feitos tanto em instrumentos, quanto no entendimento de seus componentes químicos e enzimáticos. Como consequência desse desenvolvimento, os custos para aplicação do método foi reduzido. Contudo, a introdução de novas enzimas, ácidos nucleicos, químicos e instrumentos, dificultaram a comparação de resultados entre diferentes laboratórios (BUSTIN et al., 2009).

RNA não pode servir como molde para PCR, portanto, o primeiro passo num ensaio de RT-qPCR é a produção de uma fita complementar de DNA (cDNA) através da atividade de uma transcriptase reversa retroviral (BUSTIN et al., 2009; FREEMAN; WALKER; VRANA, 1999). Este é um passo crítico para uma quantificação sensível e acurada e a quantidade de cDNA produzida deve representar a quantidade inicial de RNA. Sendo assim, a sensibilidade e especificidade da enzima utilizada deve ser criteriosa para garantir o sucesso do ensaio (WONG; MEDRANO, 2005).

Um *primer* de nucleotídeos é necessário para a síntese de cDNA, podendo ser desenhado para gene específicos ou não-específicos com vantagens e desvantagens para ambos os métodos. *Primers* hexâmeros randômicos, por exemplo, contém todas as combinações possíveis de oligonucleotídeos de seis bases e ligam-se à todo e qualquer RNA presente em amostra (FREEMAN; WALKER; VRANA, 1999).

## **1.2 Funcionamento da qPCR**

A qPCR pode ser dividida em três fases: a fase linear basal, a fase log ou exponencial e a fase platô (TICHOPAD, 2003). Durante a fase basal a qPCR está apenas em seu princípio, que vai de dez a quinze ciclos, e a emissão de fluorescência a cada ciclo ainda não é o suficiente para ser diferenciada do

ruído de fundo. A fluorescência passa a ser contabilizada durante a fase exponencial, quando atinge um limite de detecção significativamente maior que os níveis de ruído experimental de fundo. O ciclo em que a detecção ocorre é denominado Ciclo de Quantificação ou, como estabelecido mais recentemente, *Quantification Cycle* (Cq) (BUSTIN et al., 2009), cujos valores são usados para calcular os resultados (HEID et al., 1996; NESTOROV et al., 2013). Durante a fase exponencial, em condições ideais, o número de cópias dobra a cada ciclo enquanto cresce em progressão geométrica. Uma vez que os reagentes críticos se esgotam, sejam *primers*, genes repórter ou dNTPs, a reação alcança a sua fase platô e o acréscimo da fluorescência não irá mais ser relevante no processamento de dados (KUBISTA, 2001 apud KUBISTA et al., 2006).

No gráfico gerado, as curvas resultantes das multiplicações dos genes alvo são separadas na fase exponencial da reação sinalizando a diferença entre a quantidade inicial dos moldes. Essa diferença é quantificada pela comparação do número de ciclos necessários para que as amostras atinjam o limite de detecção (Cq) (KUBISTA et al., 2006).

### 1.3 Tipos de Quantificação

Dois tipos de estratégias de quantificação podem ser feitos em RT-qPCR, podendo os níveis de expressão serem avaliados numa quantificação absoluta ou relativa. A quantificação absoluta relaciona o sinal de fluorescência da PCR com o número de cópias de DNA fazendo uso de uma curva de calibração (BUSTIN, 2000; PFAFFL, 2004). As curvas de calibração podem ser realizadas a partir de DNA, RNA ou plasmídeos linearizados. A fidelidade de uma qPCR absoluta é dependente da eficiência idêntica da amplificação de ambos, calibrador e amostra. O termo quantificação “absoluta” então, pode ser interpretado erroneamente, uma vez que a quantificação é relativa à curva de calibração (BUSTIN et al., 2005; PFAFFL, 2004).

Já na quantificação relativa, as mudanças de expressão de gene nas amostras são medidas baseadas num padrão externo ou numa amostra de referência, conhecida como calibrador (LIVAK; SCHMITTGEN, 2001). Seu objetivo é comparar quantidades relativas de ácidos nucleicos entre amostras e não determinar números absolutos. No método  $\Delta Cq$ , ambos os valores de gene de referência e do gene alvo são comparados. A diferença entre seus

valores de detecção  $C_q$  ( $\Delta C_q$ ) são usados para calcular o conteúdo das amostras (CHAOUACHI; BÉRARD; SAÏD, 2013; MULLIS; FERRE; GIBBS, 1994). Existem inúmeros modelos matemáticos disponíveis para calcular a média dos genes normalizados de ensaios de quantificação relativa. Dependendo do método empregado, os modelos podem fornecer diferentes resultados (BUSTIN et al., 2009).

#### **1.4 Informações Mínimas para Publicação de Experimentos de PCR quantitativo em Tempo Real (*The Minimum Information for Publication of Quantitative Real-Time PCR Experiments*)**

Para obter dados consistentes e biologicamente relevantes de qPCR, os pesquisadores devem completar etapas técnicas complexas, aplicar uma série de controles de qualidade, usar instrumentação apropriada, gerar amplificações acuradas, e selecionar uma abordagem estatística relevante para a análise de dados. Os detalhes experimentais também precisam ser reportados de maneira transparente de modo a tornar possível a replicação dos experimentos com qualidade dos resultados de qPCR (BUSTIN et al., 2013).

O MIQE (*The Minimum Information for Publication of Quantitative Real-Time PCR Experiments*) tem por objetivo assegurar a confiança dos resultados para ajudar a garantir a integridade da literatura científica, promover consistência entre laboratórios, e aumentar a transparência nos experimentos. O MIQE é um guia que inclui as informações mínimas necessárias para o entendimento e avaliação de experimentos envolvendo qPCR. Inclui tabelas e checklists que promovem informações sobre todas as condições experimentais relevantes e características do ensaio que pesquisadores podem usar para avaliar a validade dos protocolos utilizados (BUSTIN et al., 2009).

Apesar do aumento de estudos que busquem a validação de genes de referência bacterianos para RT-qPCR desde 2009 (ROCHA; SANTOS; PACHECO, 2015), alguns laboratórios argumentam que o MIQE é apenas um guia e não precisa necessariamente ser seguido. Mesmo que seja preciso planejamento e descrição criteriosos, os benefícios de se seguir o *guideline* auxiliam na robustez dos resultados, confiança e reprodutibilidade de estudos de expressão gênica (TAYLOR; MRKUSICH, 2014).

## 1.5 Genes de Referência

Genes de referência são um controle interno da reação que possuem uma sequência diferente da sequência alvo e, para servir como um controle confiável, devem atender a importantes critérios (CHERVONEVA et al., 2010). O mais importante é que a expressão do gene de referência não deve ser alterada por diferentes condições experimentais. Ainda, deve apresentar mínima variação entre tecidos e estados fisiológicos do organismo. Na prática, isso é raro, e comumente introduzem viés dentro de um experimento, levando a erros na interpretação dos resultados (BOULTER et al., 2016). Ao que parece, os genes que preenchem perfeitamente essas condições são os que correspondem ao metabolismo básico, ou seja, que são essenciais para a sobrevivência celular (KOZERA; RAPACZ, 2013; MOSSBERG et al., 2016; THELLIN et al., 1999). Contudo, é esperado que genes que se apresentem como referência em um organismo, talvez não se expressem da mesma forma em outro. Ou ainda se comportem de maneira diferente sob uma condição experimental específica. A escolha dos genes de referência exige uma abordagem complexa para cada experimento, assim como sua validação como tal (KOZERA; RAPACZ, 2013).

A quantificação correta de um verdadeiro gene de referência permite a normalização de diferenças entre a quantidade de RNA ou cDNA amplificáveis em amostras individuais geradas por diferentes quantidades iniciais de material, qualidade do material inicial, diferenças entre preparação de RNA e síntese de cDNA, desde que o controle interno seja submetido aos mesmos passos que o gene de interesse (MOSSBERG et al., 2016; RADONIC et al., 2004).

Após a publicação do MIQE *guidelines* em 2009 até o ano de 2015, houve um aumento de publicações que se preocupassem em testar e validar genes de referência bacterianos em RT-qPCR, assim como o número de citações referentes a estes estudos, e foi possível identificar os genes mais testados e mais validados. De 17 estudos, apenas 4 conseguiram validar o gene *16S*, sinalizando que talvez não seja uma boa opção para uma abordagem inicial como foi descrito em trabalhos anteriores (CARVALHO et al., 2014; LIU et al., 2013; ROCHA; SANTOS; PACHECO, 2015; TURRONI et al., 2011). A maior

proporção de genes bacterianos validados nesse período pertenciam a rotas metabólicas basais como replicação de DNA e metabolismo de RNA.

Quando comparados a dados de expressão gênica no banco de dados *Pathosystems Resource Integration Center* ([www.patricbrc.org](http://www.patricbrc.org)) (WATTAM et al., 2017), uma lista de genes para uma seleção inicial de genes candidatos a referência para RT-qPCR em diferentes espécies bacterianas foi obtida (ROCHA et al., 2015). Com mais de 90% de estudos, em diferentes filos bacterianos, comprovando estabilidade de expressão, o topo da lista era composto, em ordem, pelos genes *gyrB* (DNA girase B), *gyrA* (DNA girase A), *era* (GTP-binding proteing ERA), *secA* (translocase subunidade SecA) e *dnaG* (DNA primase). Os dados corroboram a ideia de que estes são bons candidatos a genes de referência, que serviriam para uma primeira abordagem em análise de expressão e, portanto, bons candidatos para a construção de um *kit* de normalização bacteriana (ROCHA; SANTOS; PACHECO, 2015).

## 1.6 Normalização

A normalização é um componente essencial de um ensaio de qPCR confiável (BUSTIN et al., 2009). Erros específicos na quantificação de transcritos de mRNA são compostos por qualquer variação na quantidade de material inicial entre as amostras. Isso é especialmente relevante quando as amostras são obtidas a partir de diferentes indivíduos, e resultarão em má interpretação dos perfis de expressão dos genes alvo. Estes erros podem ser corrigidos pela normalização (KOZERA; RAPACZ, 2013; NESTOROV et al., 2013a; THELLIN et al., 1999).

Não existe uma estratégia de normalização universalmente aceita, assim como não existe procedimento livre de erros. A melhor forma de se assegurar a precisão, confiança e reprodutibilidade é equalizar as amostras. Garantir o mínimo de similaridade na quantidade de RNA para síntese de cDNA por transcrição reversa e a medição de um controle interno como um gene de referência (HUGGETT; O'GRADY; BUSTIN, 2015; NESTOROV et al., 2013). Para os genes serem utilizados como normalizadores, devem ter sido validados experimentalmente para o organismo e condições dentro do desenho experimental de um estudo (BUSTIN et al., 2009; TAYLOR; MRKUSICH, 2014).

A normalização contra um único gene de referência não é aceitável. A exceção seria por evidência concreta que confirme a manutenção estável de sua expressão dentro das condições de estudo. O melhor número de genes de referência pode ser determinado ainda pela abordagem algorítmica utilizada para a validação dos mesmos (ANDERSEN; JENSEN; ORNTOFT, 2004; VANDESOMPELE et al., 2002).

O gene de governança ideal deve ser expresso em um nível constante em diferentes tecidos de um organismo, em todos os estágios do desenvolvimento, e não deve ser afetado pelas condições experimentais. Deverá também ser expresso na mesma quantidade que o RNA alvo do estudo (THELLIN *et al*, 1999 apud KOZERA; RAPACZ, 2013). A escolha no nível aceitável de variação vai depender no grau de resolução requerida. Mesmo se o gene escolhido for variável talvez isso não importe, desde que a diferença entre os genes alvo seja maior que entre os genes de referência (HUGGETT et al., 2005).

Existem uma série de ferramentas de análise baseadas na plataforma excel que permitem a avaliação de múltiplos genes de referência. O geNorm permite o escalonamento dos melhores genes escolhidos através da média geométrica dos valores de expressão dos cDNA candidatos. O Bestkeeper também determina os melhores genes por cálculo da média geométrica, mas só trabalha com dados brutos ao invés de dados convertidos para números de cópias e, como o geNorm, também é gratuito. O NormFinder, por sua vez, não apenas avalia a variação como também escalona os genes em quanto eles diferem entre diferentes grupos (HUGGETT et al., 2005). Esta ferramenta também é a menos afetada por co-regulação de genes (ANDERSEN; JENSEN; ORNTOFT, 2004).

## 2. JUSTIFICATIVA

O desenvolvimento da PCR quantitativa em tempo real foi um marco para comunidade científica, com sua capacidade de detectar e mensurar quantidades ínfimas de ácidos nucleicos numa diversidade de amostras de todas as fontes, e por sua praticidade e excelência permitindo o uso em diversas áreas do conhecimento; das ciências da vida, da agricultura e medicina (BUSTIN, 2000; KUBISTA et al., 2006). Sua explosão em popularidade, no entanto, se refletiu na diversidade em protocolos para diferentes etapas dentro da metodologia do RT-qPCR, não existindo então uma padronização da técnica que se faz ainda mais necessária quando se trata de suas aplicações clínicas. Outro fator importante e que, quando feito erroneamente, pode prejudicar um trabalho é o uso de genes de referência não validados e abordagens estatísticas diversas e muitas vezes inadequadas (BUSTIN et al., 2009).

Num estudo prévio, (ROCHA; SANTOS; PACHECO, 2015) nosso grupo de pesquisa avaliou a literatura na busca por trabalhos que envolvessem a seleção e validação de genes de referência bacterianos em ensaios de RT-qPCR. O estudo foi capaz de determinar as melhores vias metabólicas para uma seleção inicial de genes, identificou os genes mais abordados, os mais validados e também os mais estáveis, baseando-se nos dados experimentais de análise de expressão no banco de dados *Pathosystems Resource Integration Center* (PATRIC). A lista de genes avaliados no estudo deve servir como guia para uma seleção inicial de candidatos a genes de referência para diferentes espécies bacterianas.

Bancos de dados como *PrimerBank* da Universidade de Massachusetts contêm milhares de *primers*, principalmente para humanos e camudongos. Estão ainda disponíveis no mercado *kits* de normalização de DNA para otimizar pesquisa e diagnóstico que utilizem a metodologia de RT-qPCR. Entretanto, apesar da grande importância médica e científica, ainda não existem *kits* para normalização em organismos bacterianos, tampouco *primers* para genes de referência que sejam viáveis não somente para uma única espécie. Uma vez que existem genes que mantêm a sua estabilidade sob diferentes condições experimentais em diversas espécies bacterianas, o desenvolvimento de um kit

de *primers* universais de genes de referência de eficiência comprovada, seria de grande utilidade para pesquisa e fácil aceitação no mercado biotecnológico, reduzindo etapas, e conseqüentemente, tempo, em pesquisas de expressão gênica fazendo uso de RT-qPCR.

### 3. OBJETIVOS

Desenvolver e validar um conjunto de *primers* universais de genes de referência bacterianos para normalização em ensaios de PCR quantitativa em tempo real.

#### 3.1 Objetivos Específicos

- Buscar por regiões conservadas para os genes *gyrA*, *gyrB*, *era*, *secA* e *dnaG* entre espécies bacterianas de filos bacterianos diferentes;
- Desenhar *primers* universais para os genes *gyrA*, *gyrB*, *era*, *secA* e *dnaG* em filos bacterianos relevantes;
- Analisar os genes de referência selecionados em diferentes espécies de bactéria crescidas sob condições diversas;
- Validar e padronizar o uso dos *primers* dos genes de referência por RT-qPCR.

#### 4. METODOLOGIA

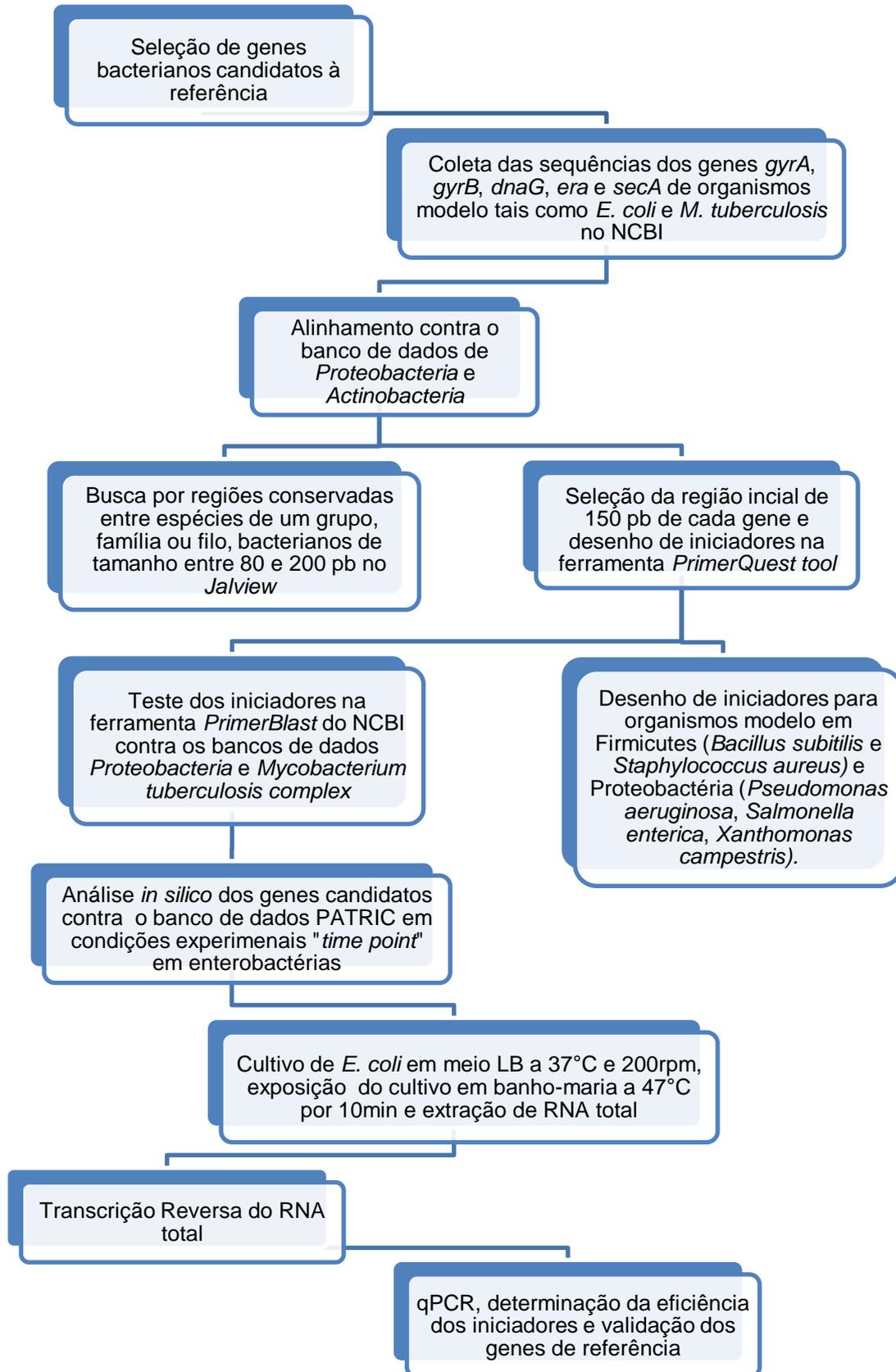


Figura 1 Fluxograma da metodologia de trabalho.

#### 4.1 Desenho de *Primers*

Após a seleção prévia dos genes candidatos (ROCHA; SANTOS; PACHECO, 2015), a etapa seguinte foi determinar os organismos cuja sequência de DNA serviria de base para as etapas posteriores de alinhamento através da ferramenta BLAST (Disponível em: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Estes organismos foram selecionados com base na sua importância clínica, na disponibilidade de seu genoma e vasta caracterização de seus genes. Inicialmente foram selecionados os organismos *Escherichia coli* e *Mycobacterium tuberculosis* pertencentes aos filos *Proteobacteria* e *Actinobacteria*, respectivamente. Com o auxílio de operadores booleanos foram realizadas pesquisas no banco de dados “Gene”, presente no NCBI (*National Center for Biotechnology Information*; Disponível em: <https://www.ncbi.nlm.nih.gov/gene/>) (Figura 1), para cada um dos genes candidatos à referência como descrito no exemplo a seguir:

- “*Escherichia coli* AND *gyrB* DNA *gyrase B*”
- “*Escherichia coli* AND *gyrA* DNA *gyrase A*”
- *Escherichia coli* AND *dnaG* DNA *primase*
- *Escherichia coli* AND *era* GTP *binding protein era*
- *Escherichia coli* AND *translocase subunit secA*

Na janela de resultados foram selecionadas as opções relacionadas aos genes e espécies em questão e, após a verificação das informações, as sequências eram salvas em formato de arquivo FASTA e devidamente identificadas (Fig 2 e 3).

Após a obtenção das sequências de DNA para *Escherichia coli* e *Mycobacterium tuberculosis* para os genes *gyrA*, *gyrB*, *dnaG*, *era* e *secA*, cada uma delas foi submetida ao alinhamento contra sequências do banco de dados do NCBI de organismos classificados nos filos *Proteobacteria* e *Actinobacteria*. Uma vez que *E. coli* e *M. tuberculosis* figuram entre os organismos mais estudados, conseqüentemente existe um número considerável de sequências depositadas, para diversas cepas e variantes de diversos países do mundo. Sendo assim, para otimizar a quantidade de espécimes diferentes obtidas, a opção para excluir todos os resultados para estas espécies foi feita na interface

da ferramenta. Ainda de modo a abranger o maior número possível de resultados, nos parâmetros do algoritmo de busca, foi selecionado a opção “blastn”.

Foram selecionadas, e salvas em formato FASTA, as sequências correspondentes em espécies e cepas que possuíam entre 75% e 100% de identidade, eliminando alinhamentos com cobertura de sequência inferior a 80% a fim de evitar a seleção de pequenos fragmentos de DNA. Logo após, as sequências foram todas organizadas, identificadas apropriadamente e submetidas ao alinhamento pela ferramenta online *ClustalOmega* do Instituto Europeu de Bioinformática (Disponível em: <http://www.ebi.ac.uk/Tools/msa/clustalo/>). Nenhum parâmetro original foi modificado além da sinalização de que as sequências se tratavam de DNA.

**Figura 2** Página de resultados do alinhamento pela ferramenta BlastN do NCBI. Como exemplificado estão selecionadas as espécies com melhor colocação em porcentagem de identidade e score no alinhamento.

Sequences producing significant alignments:

Select: All None Selected: 45

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis strain SP38, complete genome</a>	3463	3508	100%	0.0	100%	<a href="#">CP015773.2</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium africanum strain UT307 genome</a>	3463	3508	100%	0.0	100%	<a href="#">CP014617.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium microti strain 12, complete genome</a>	3463	3508	100%	0.0	100%	<a href="#">CP010333.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis strain 1595, complete genome</a>	3463	3508	100%	0.0	100%	<a href="#">CP012095.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis subsp. bovis AF2122/97, complete genome</a>	3463	3508	100%	0.0	100%	<a href="#">BX248333.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis strain ATCC BAA-935, complete genome</a>	3463	3508	100%	0.0	100%	<a href="#">CP009449.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. Tokyo 172 substrain TRCS, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP014566.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis strain BCG-1 (Russia), complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP013741.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis strain BCG-1 genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP011455.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis BCG strain Russia 368, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP009243.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis BCG strain 3281, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP008744.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. ATCC 35743, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP003494.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. Korea 1168P, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP003909.2</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. Mexico, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">CP002095.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. Moreau RDJ, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">AM412059.2</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium africanum GM041182, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">FR878060.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis BCG str. Tokyo 172 DNA, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">AP010918.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium bovis BCG Pasteur 1173P2, complete genome</a>	3458	3502	100%	0.0	99%	<a href="#">AM408590.1</a>
<input type="checkbox"/>	<a href="#">Mycobacterium bovis strain 30, complete genome</a>	3454	3499	100%	0.0	99%	<a href="#">CP010332.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mycobacterium canettii CIFT 140060008, complete genome</a>	3454	3499	100%	0.0	99%	<a href="#">FQ203507.1</a>

Os resultados dos alinhamentos foram salvos e abertos no aplicativo para *desktop Jalview*. O *Jalview* é um programa gratuito para edição, visualização e análise de alinhamentos múltiplos de sequências (WATERHOUSE et al., 2009). Para cada um dos genes e filios bacterianos foram procuradas regiões de homogeneidade entre 80 e 200 pb (pares de bases) para serem usadas na construção dos *primers forward* e *reverse* e obter um produto amplificado dentro dos parâmetros apropriados à metodologia de qPCR e no máximo 150

pares de base. Espécies que se distanciavam muito da semelhança eram excluídas na edição do arquivo, permitindo a melhoria da visualização dessas regiões como consequência do agrupamento de proximidade filogenética.

Foi decidida a adoção de uma segunda estratégia com base nos resultados dos alinhamentos anteriores de *gyrA* e *gyrB*, sendo selecionados os primeiros 150 pares de bases da sequência de cada gene de *Escherichia coli* e *Mycobacterium tuberculosis* e submetidas individualmente à ferramenta de desenho de *primers* *PrimerQuest Tool* da IDT<sup>®</sup> *Integrated DNA Technologies* (Disponível em: <https://www.idtdna.com/Primerquest/Home/Index>), em que era requerido o desenho de dois *primers* para experimentos com uso de corantes intercalantes de DNA. A ferramenta entregou os números de *primers* ideais possíveis dentro da sequência inserida.

Após a obtenção desses resultados, cada par de *primer* sugerido para cada gene foi, por sua vez, submetido à uma PCR digital, utilizando a ferramenta *Primer-Blast* do NCBI (Disponível em: <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>), de modo a avaliar as espécies para quais os *primers* amplificariam a sequência correta. Para *E. coli*, a busca foi realizada dentro do filo *Proteobacteria* e, para *M. tuberculosis*, dentro do complexo *Mycobacterium tuberculosis*. Todos os resultados de espécies e cepas que possuíam mau-pareamento de até dois pares de bases próximo à terminação 3' foram descartados, assim como ampliações inespecíficas identificadas pela diferença no tamanho esperado de *amplicon*.

#### **4.2 Análise *in silico* da estabilidade dos genes candidatos a genes de referência**

Uma segunda análise dos dados no banco de dados do *Pathosystems Resource Integration Center* (<https://www.patricbrc.org>) foi feita para cada um dos 5 genes em organismos modelo e de interesse clínico como: *Bacillus subtilis*, *Escherichia coli*, *Burkholderia pseudomallei*, *Pseudomonas aeruginosa*, *Streptococcus mutans*, *Staphylococcus aureus* e *Helicobacter pylori*. Todos os dados de transcriptoma foram baixados em forma de tabela e, por análise em *Excel*<sup>®</sup>, mantidos aqueles que correspondiam à condição experimental “*time point*”, ou seja, cujo estudo avaliasse a estabilidade gênica em diferentes

momentos de um cultivo bacteriano. Eles foram separados em grupos correspondentes a fase de cultivo a que pertenciam, como fase exponencial inicial ou tardia e fase estacionária inicial ou tardia. Qualquer outro dado não correspondente à esses parâmetros foram excluídos de análises posteriores.

Os dados de estabilidade para cada condição dos genes foram submetidos à análise estatística teste T de Student, por meio do *Excel*<sup>®</sup> para verificar se havia diferença estatística entre as condições. Caso houvesse diferença, então o gene não seria estável dentro do desenho experimental e os *primers* desenhados para o mesmo não seriam sintetizados. Como consequência, o mesmo não seria incluso nas demais análises.

#### **4.3 Primers de para organismos modelo em *Proteobacteria* e *Firmicutes***

O desenho de *primers* para os genes *gyrB*, *dnaG*, *secA* e *16S* foi expandido para organismos-modelo bacterianos cujos dados de transcriptoma se encontram disponíveis na plataforma *Pathosystems Resource Integration Center* (PATRIC). Foram escolhidos dentro do filo *Firmicutes*: *Bacillus subtilis* e *Staphylococcus aureus*. Para *Proteobacteria*: *Pseudomonas aeruginosa*, *Salmonella enterica*, *Xanthomonas campestris*.

A estratégia adotada para o desenho desses *primers*, foi buscar diretamente a sequência no banco de dados do NCBI como descrito no ítem 4.1, e a partir dela, obter diretamente os *primers* através do *PrimerQuest Tool*. Cada opção de *primers* para cada organismo teve sua amplificação testada por meio de PCR virtual, se utilizando da ferramenta *PrimerBlast* do NCBI. A opção que amplificasse para o maior número de espécies e que não possuísse amplificações inespecíficas, era mantida e as demais desconsideradas.

O gene *gyrA* foi desconsiderado por ter performance nos critérios de conservação inferior ao *gyrB* e por serem co-regulados. Isso significa que juntos no experimento poderiam influenciar a análise estatística e levar a interpretações equivocadas a respeito de suas estabilidades (ROBLEDO et al., 2014).

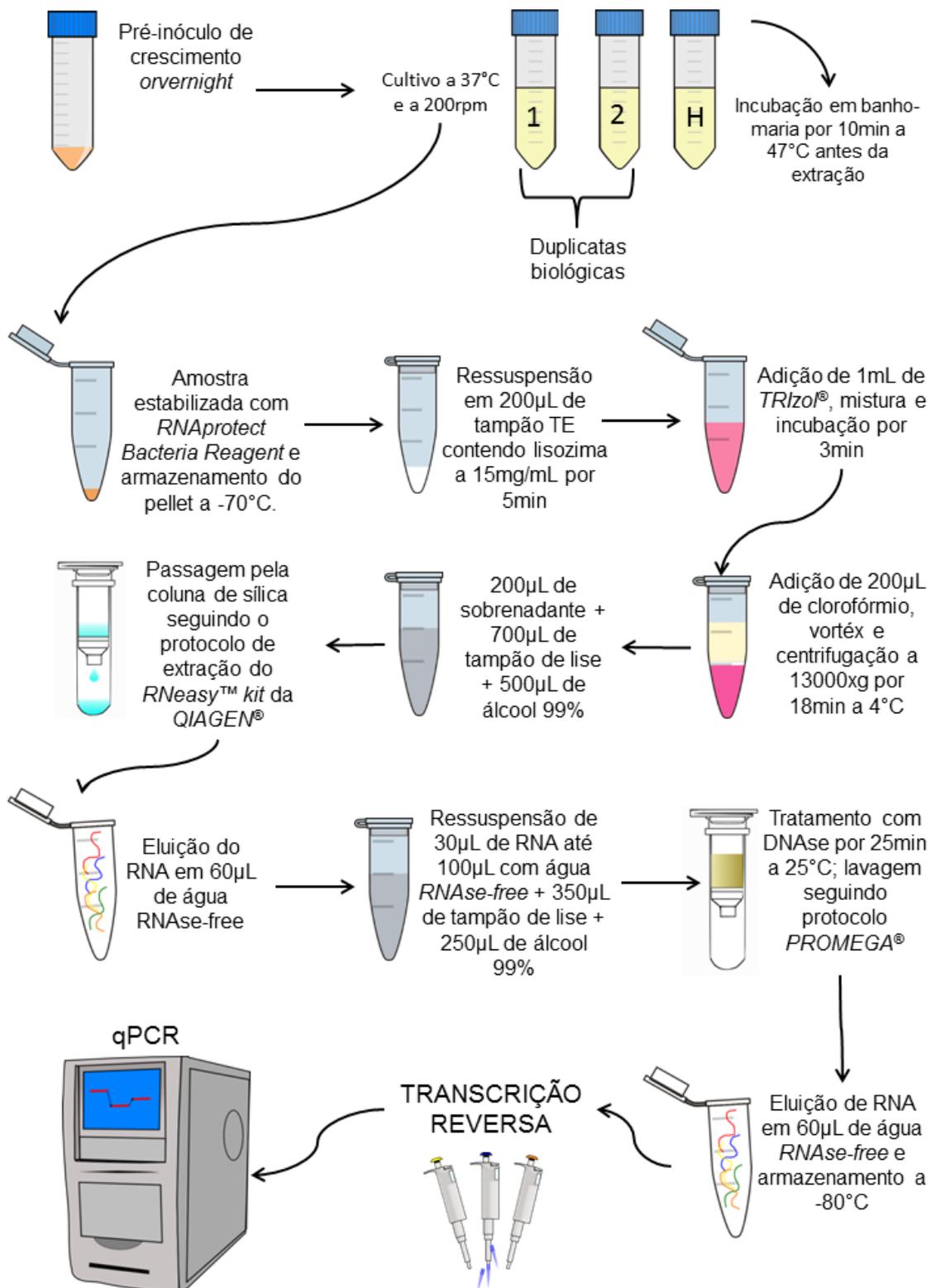
Já o gene *16S* é comumente utilizado na literatura como normalizador e a comparação do seu desempenho neste estudo com os demais genes se faz importante (CARVALHO et al., 2014; LIU et al., 2013).

#### 4.4 Cultivo Bacteriano

Uma cepa de *Escherichia coli* BL21 foi cultivada em meio Luria-Bertani (LB). O pré-inóculo foi realizado em fluxo laminar, com 50µL de células suspensas em glicerol a 30% em 5mL de meio LB num tubo *Falcon* de 50mL, junto ao controle negativo, e incubado *overnight* (por 16 horas) com agitação a 200rpm, 37°C. Na manhã do dia seguinte, foi realizado o inóculo com diluição 1:50, correspondendo à 100µL de cultura em 5mL de meio em três réplicas. O crescimento foi monitorado por meio de espectrofotometria até que a  $DO_{600nm}$  (densidade óptica) atingisse a fase logarítmica de crescimento, entre 0,5 e 0,7. Previamente à extração de RNA, uma das culturas foi mantida em banho-maria por 10min a 47°C para indução do estresse térmico. O mesmo protocolo de obtenção de amostra foi feito para cada cultura na fase exponencial tardia/estacionária do crescimento, com  $DO_{600nm}$  acima de 2,0.

## 4.5 Extração de RNA

Figura 3 Representação esquemática do método de extração de RNA bacteriano utilizado.



A extração de RNA foi otimizada e feita pela combinação de diferentes kits comerciais, sendo eles. O *Trizol*<sup>TM</sup> da *Ambion*<sup>®</sup> o *SV Total RNA Isolation System* da *Promega*<sup>®</sup> e o *RNeasy Protect Bacteria*<sup>TM</sup> da *QIAGEN*<sup>®</sup>. Quinhentos microlitros de cada cultura (duas replicatas e o estresse térmico) foram transferidos para microtubos de 2mL e imediatamente misturados a 1mL do reagente *RNAprotect Bacteria* (*QIAGEN*<sup>®</sup>) por vortéx e deixados em temperatura ambiente por 5min para então serem centrifugados por 10min a 10.000xG e o sobrenadante removido por inversão.

Ao pellet de cada cultura foram adicionados 200µL de tampão Tris-EDTA (30mM Tris-Cl, 1mM EDTA, pH=8,0) contendo 15mg/mL de lisozima, misturados por vortéx e incubados por 5min a temperatura ambiente. Foram então adicionados 1mL de *TRIZOL*<sup>TM</sup> a cada microtubo, misturados por vortéx e deixados em temperatura ambiente por mais 5min. Logo após, foram adicionados 200µL de clorofórmio a cada amostra, misturados por vortéx vigorosamente e submetidas a centrifugação por 18min, 13000xg a 4°C. Para cada amostra, foram retirados 200µL de sobrenadante e transferidos para outro microtubo, onde foram adicionados 700µL de tampão de lise do kit *QIAGEN*<sup>®</sup> e 500µL de álcool 99% e a mistura então posta para passar pela membrana do kit *Rneasy*<sup>TM</sup> mini kit da *QIAGEN*<sup>®</sup> por centrifugação a 8000xg por 30s. Esta etapa foi realizada mais uma vez aproveitando tanto o volume da mistura como o sobrenadante da centrifugação com *Trizol*<sup>TM</sup>. Em seguida foi adicionado à membrana 700µL de tampão de lavagem e mais uma vez submetida a centrifugação a 8000xg por 30s. Mais duas lavagens foram feitas com 500µL de tampão RPE também a 8000xg por 30s e ressuspensão feita, por duas vezes, com 30µL de água *RNAse-free* na velocidade máxima da centrífuga por 2min.

A mensuração da concentração de RNA total foi realizada com o uso da *µDrop Plate*<sup>TM</sup> da *Thermo Scientific*<sup>®</sup>. Aplicou-se, em duplicata para cada extração, 2µL de amostra por poço, 2µL de água como controle negativo, e depois submetendo a leitura por espectrofotometria entre 260 e 280nm de comprimento de onda.

#### 4.5.1 RNA Cleanup

Do volume total de 60µL de amostra de RNA obtido na extração, foram usados a metade, 30µL, para re-purificação pela mistura dos protocolos de *cleanup* da *QIAGEN*<sup>®</sup> e da DNaseI da *PROMEGA*<sup>®</sup>. Esses 30µL foram ressuspensos para o volume de 100µL com água *RNAse-free* misturado, por pipetagem, a 350µL de tampão de lise e 250µL de álcool 99%. Todo o volume (cerca de 700µL) foi posto sobre a coluna para centrifugação a 8000xG por 30s.

Para cada amostra de RNA em repurificação, foram preparados o mix de incubação de DNase (*PROMEGA*<sup>®</sup>), combinando 40 µL de tampão, 5 µL de MnCl<sub>2</sub> a 0,09M e 5µL de DNaseI, que foram aplicados diretamente sobre a membrana das colunas. Após uma incubação por 25 minutos entre 20-25°C foi adicionado 200µL de solução de parada da reação de DNase em cada coluna e centrifugadas por 1 minuto a 13000xg, seguido de mais três passos: Adição de 600µL de solução de lavagem de RNA e centrifugação por 1 min a 13000xg, descarte do material residual, e mais uma adição de 250µL da mesma solução, seguido de centrifugação por 2 min em velocidade máxima.

Na última etapa, a coluna foi transferida para dentro de outro microtubo, e foram adicionados 60µL de água *RNAse-free* sobre a membrana para eluir o RNA numa última centrifugação por 1 min a 13000xg. Os resultados da repurificação foram mensurados novamente com o uso da *µDrop Plate*<sup>™</sup> da *Thermo Scientific*<sup>®</sup>.

### 4.6 Transcriptase Reversa seguida de PCR Quantitativa (RT-qPCR)

#### 4.6.1 Transcrição Reversa

A transcrição reversa foi realizada fazendo uso do kit comercial *GoScript*<sup>™</sup> *Reverse Transcription System* (Promega) seguindo o protocolo do fabricante, adicionando apenas uma modificação, que foi dobrar o volume das reações para minimizar erros de manipulação por pipetagem. Para cada amostra foram utilizadas 200ng de RNA e misturadas em seus respectivos microtubos a 1µg de *primers* randômicos e água *RNAse-free* até o volume de 10µL. Então, os microtubos foram postos em um termociclador onde foram

aquecidos a 70°C por 5min e depois mantidos refrigerados a 4°C até serem usados novamente.

Em paralelo foram montados os mixes de transcrição reversa pela adição, por reação, de 8µL de tampão de reação *GoScript<sup>TM</sup> 5x*, 3µL de MgCl<sub>2</sub>, 2µL de mix de nucleotídeos, 1µL de inibidor de nuclease *RNasin<sup>®</sup>*, 2µL de transcriptase reversa e água RNase-free até o volume de 30µL.

Ambas as reações foram combinadas em um único microtubo e postas novamente no termocilador a 25°C por 5min para anelamento dos *primers*, seguido de 42°C por 1h para extensão e 15min a 70°C para inativação da transcriptase reversa. As amostras de DNA complementar (cDNA) sintetizadas foram então armazenadas a -20°C até serem utilizadas novamente.

#### 4.6.2 qPCR

Para a realização da PCR quantitativa foi utilizado o kit comercial *GoTaq<sup>®</sup> qPCR Master Mix* da *Promega<sup>®</sup>*. O teste foi todo montado em duplicata, com controle negativo, amostra de cDNA1 e cDNA2 (como foram denominadas as duplicatas biológicas), e cDNAH (como foi chamado a amostra para a condição de estresse térmico ou *Heat Shock*). Para cada amostra foram testados, na mesma placa, os genes *gyrB* (foram testados dois pares de *primers* diferentes), *dnaG*, *secA* e *16S*. Para cada gene também foram realizadas curvas de diluição de amostra de 1x, 10x, 100x, 1000x e 10000x para verificar a eficiência dos *primers*. As amostras de RNA, por sua vez, também foram incluídas previamente em outra análise, nas mesmas condições e diluições realizadas com os cDNA correspondentes, afim de averiguar a contaminação por DNA genômico (gDNA).

**Tabela 1** Parâmetros de ciclagem de PCR quantitativa como sugerido pelo fabricante utilizados na reação pro termociclador.

Componentes	Ciclos	Temperatura	Tempo
Ativação Hot-Start	1	95° C	2min
Desnaturação	44	95° C	15s
Anelamento/Extensão		60° C	1min
Extensão final	1	72°C	5min
Desassociação	1	60~95° C	2 min

Cada reação individual deveria conter 10µL de *GoTaq<sup>®</sup> qPCR Master Mix*, 0,2µM de *primers forward* e 0,2µM *reverse* do gene a ser testado, 5,2µL de água *RNAse-free*, resultando num volume final de 16µL. Em microtubos foram montados os mixes de reação para cada par de *primer* dos genes testados para então serem aplicados em suas respectivas posições na placa de 96 poços. Então adicionou-se 4µL de cDNA nas posições correspondentes e a placa foi selada com adesivo ótico em seguida. O termociclador utilizado foi o *CFX96 Touch<sup>TM</sup> Real-Time PCR Detection System (BioRad<sup>TM</sup>)*, onde foi selecionado *SYBR<sup>®</sup>* como corante de detecção para toda placa e, programando um total de 44 ciclagens na máquina *CFX96 Touch<sup>TM</sup> Real-Time PCR Detection System*, cujos resultados foram obtidos e analisados por meio do *CFX Manager<sup>TM</sup> Software da Bio-Rad*.

#### **4.6.3 Determinação da eficiência dos *primers***

A eficiência de amplificação para os *primers* utilizados foi avaliada a partir de diluições seriadas de cDNA de até 5 pontos (1x, 10x, 100x, 1000x e 10000x), e as reações realizadas como descrito no item 4.5.2. Os ciclos de quantificação (Cq) para cada gene foram utilizados para gerar a curva padrão de cada par de *primer* testado, sendo calculados os coeficientes de correlação entre os pontos e a reta plotada para validação da regressão, com o emprego do *software Excel<sup>®</sup> 2013*. A eficiência da amplificação (E) foi obtida através da ferramenta online “*qPCR Efficiency Calculator*” *ThermoFisher Scientific<sup>®</sup>* que faz uso da equação  $E = 10^{(-1/\text{slope})}$  (Rasmussen, 2001), em que o *slope* equivale à inclinação da reta obtida da regressão entre os valores de Cq dos transcritos e os valores do logaritímicos das diferentes diluições de cDNA.

#### **4.6.4 Verificação da quantidade de DNA genômico**

A quantidade de DNA genômico foi verificada por meio de qPCR, em um teste distinto da estabilidade dos genes testados e eficiência dos *primers*. Comparou-se os dados de Cq entre as amostras cDNA e seu RNA correspondente. Para isso, o volume correspondente a 80ng (o mesmo usado pra sintetizar o cDNA neste teste) foi diluído em água *RNAse-free* até o volume de 20µL, que corresponde a reação de cDNA. Então, 4µL dessa diluição foi

usado como template para a reação de qPCR. Os dados foram analisados via *CFX Manager™ Software da Bio-Rad*.

#### **4.6.5 Avaliação da estabilidade dos genes candidatos a referência**

Para verificar a estabilidade dos genes candidatos a referência, foram comparadas as alterações nos níveis de expressão entre as amostras obtidas em condição normal de crescimento em relação à condição de choque térmico (*heat shock*) por 10min a 47°C. Os valores de Cq obtidos para os 4 genes foram convertidos em quantidades relativas, por meio de correção dos valores de eficiência de amplificação para cada par de *primers*, fazendo uso da fórmula descrita por  $Q = E^{\Delta Ct}$  (GALISA et al., 2012), onde E refere-se ao valor de eficiência de amplificação do gene, e o  $\Delta Ct$  corresponde a diferença entre o valor da amostra em questão e o menor valor de Ct dentre todos o conjunto de valores.

O *NormFinder* (LIU et al., 2013) utiliza um algoritmo para identificar o gene normalizador otimizado entre um conjunto de genes, através do *ranking* desses genes de acordo com a sua variação de expressão inter- e intragrupo (S). O gene de referência mais estável é aquele que obtiver o menor valor de S (LIU et al., 2013).

Também foi utilizada a ferramenta online *RefFinder* (XIE et al., 2011) (disponível em: <http://150.216.56.64/referencegene.php>) que, por meio de dados brutos de Ct, faz uso do método  $\Delta$ -CT comparativo e dos algoritmos do NormFinder, geNorm, BestKeeper para comparar e apresentar em forma de tabelas e gráficos os genes mais estáveis entre as mais usadas metodologias de normalização (XIE et al., 2011).

## 5. RESULTADOS

### 5.1 Desenho de *Primers*

#### 5.1.1 *Primers* de Proteobactérias

Na primeira estratégia de desenho de *primers*, as sequências dos genes *dnaG*, *era*, *gyrA*, *gyrB*, *secA* de *E. coli* foram submetidas ao alinhamento no *BLASTn* contra o banco de dados do filo *Proteobacteria*. As primeiras posições de homologia, quando não pertencentes a própria *Escherichia coli*, eram ocupadas predominantemente por organismos membros da família taxonômica *Enterobacteriaceae* em função de sua proximidade filogenética e grande quantidade de sequências depositadas.

As sequências de cada um dos genes que foram selecionadas e alinhadas, apresentaram clara divisão entre grupos bacterianos distintos. Na visualização pelo Jalview, para o gene *gyrB*, no agrupamento em que se encaixou *E. coli*, apresentou-se o maior nível de conservação gênica, principalmente na região inicial do gene com cerca de 100 pares de base (Figura 4). Com a exclusão dos grupos com menos homologia (as espécies: *Morganella morganii*, *Yersinia rohdei*, *Yersinia enterocolitica*, *Routella ormithinolytica*, *Klebsiella oxytoca*, *Enterobacter aerogenes*, *Klebsiella pneumoniae*, *Citrobacter freundii*, *Salmonella enterica*, *Shigella dysenteriae*, *Serratia rubidaea*, *Chania multiturdinenses*, *Serratia fronticola*, *Proteus mirabilis*, *Candidatus sodalis*, *Yersinia pestis*, *Yersinia ruckeri*, *Pantoea sp.*, *Pectobacterium carvorum*, *Edwardsiella tarda*, *Rahnella aquatis*, *Serratia marcescens*, *Serratia plymuthica*, *Serratia liquefaciens*), a porcentagem de homologia entre as sequências aumentou significativamente (Figura 5). O gene *gyrA*, por sua vez, demonstrou um padrão semelhante com a região inicial com cerca de 160 pares de bases, tendo bastante conservação entre espécies. Requerendo, portanto, os mesmos ajustes que *gyrB*, minimizando *mismatches* (Figuras 6 e 7).

Quando submetidos ao *PrimerQuest Tool*, *gyrA* e *gyrB* foram capazes de gerar ao menos um par de *primers* que não possuíam nenhum *mismatch* com as espécies inclusas na imagem final produzida (*gyrB*: *Raoutella ormithinolytica*, *Enterobacter lignolyticus*, *Enterobacter coaccae*, *Klebsiella variicola*, *Citrobacter*

rodentium, *Escherichia coli*, *Klebsiella pneumoniae*, *Citrobacter freundii*, *Shigella sonnei*, *Shigella flexneri*, *Enterobacter asburiae*, *Enterobacter coaccae*, *Salmonella enterica*) (*gyrA*: *Klebsiella pneumoniae*, *Klebsiella variicola*, *Enterobacter lignolytica*, *Citrobacter freundii*, *Enterobacter coaccae*, *Salmonella enterica*, *Shigella sonnei*). Para o gene *secA*, os ajustes para se obter uma região minimamente aceitável reduziram seu grupo de organismos a apenas 6 indivíduos (*Escherichia coli*, *Shigella flexneri*, *Citrobacter freundii*, *Klebsiella oxytoca*, *Enterobacter asburiae*), produzindo um único par de *primers* viável que possuiria *mismatches* quando usados em outras espécies que não *E. coli* (Figura 12 e 13). Os genes *dnaG* e *era*, por sua vez, mesmo com a divisão de grupos pelo alinhamento, não apresentaram conservação suficiente para viabilizar o desenho de *primers* universais sem *mismatches* que não compromettessem sua eficiência (Figuras 8, 9, 10 e 11).

**Figura 4 Visualização parcial do alinhamento de seqüências para o gene *gyrB* via Jalview.** Os tons azul, do mais claro ao mais escuro, representam o acréscimo da porcentagem de identidade. É possível identificar um grupo de bactérias cujo gene se encontra em maior conservação que outro.

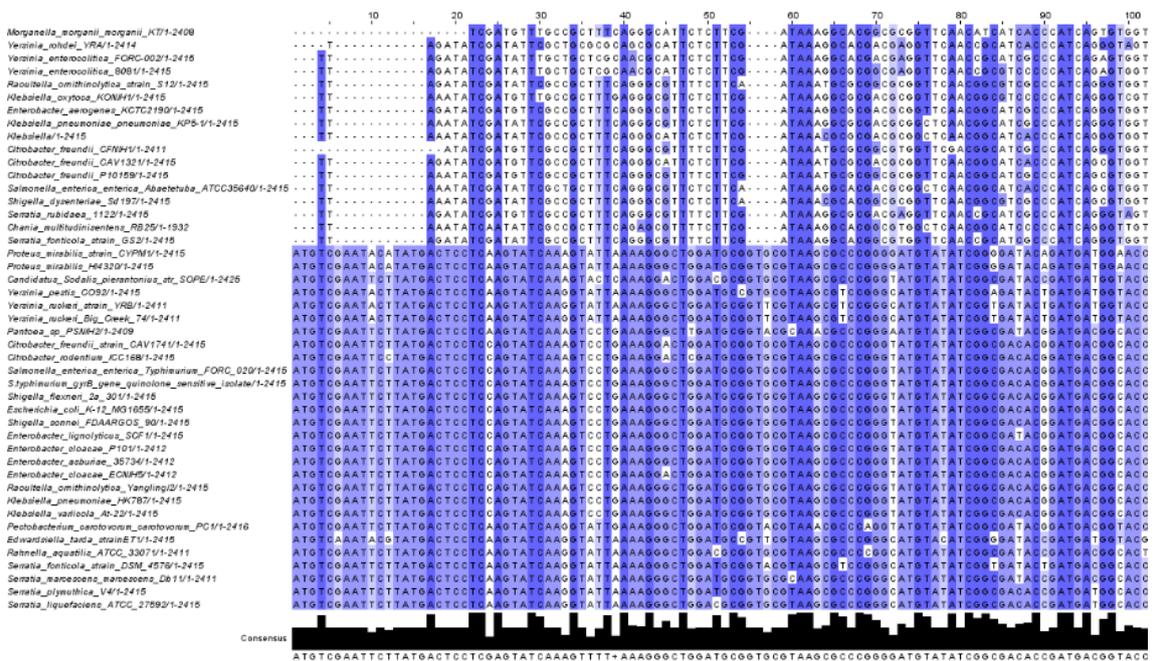


Figura 5 Visualização de uma região conservada do alinhamento de seqüências para o gene *gyrB*. A cor azul escuro representa a porcentagem de semelhança. Marcado em cinza, está a seqüência do gene *gyrB* de *Escherichia coli* utilizada como molde para o alinhamento.

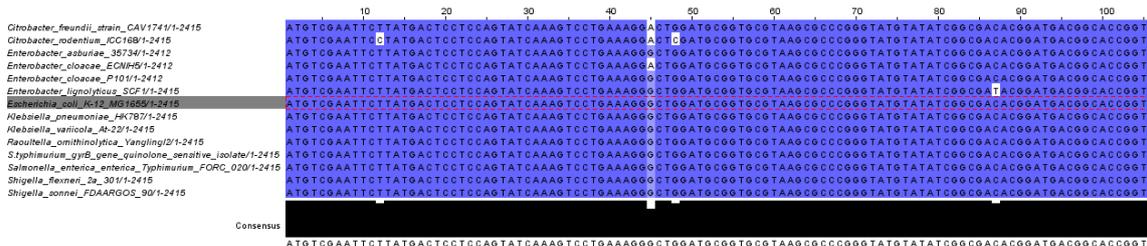


Figura 6 Visualização parcial do alinhamento de seqüências para o gene *gyrA* via Jalview. Os tons azul, do mais claro ao mais escuro, representam o acréscimo da porcentagem de identidade.

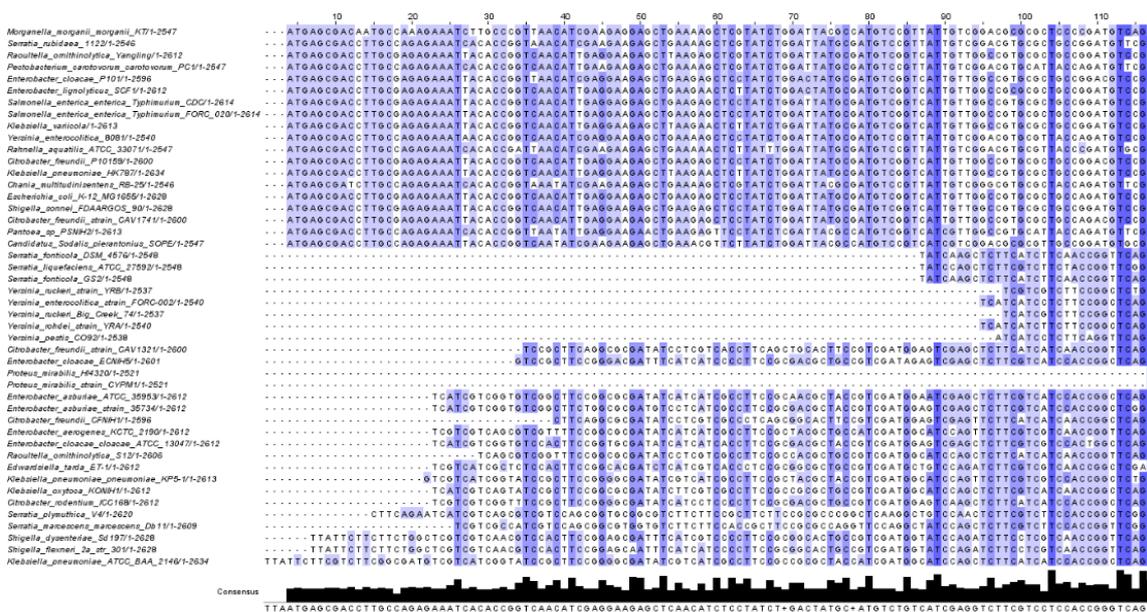
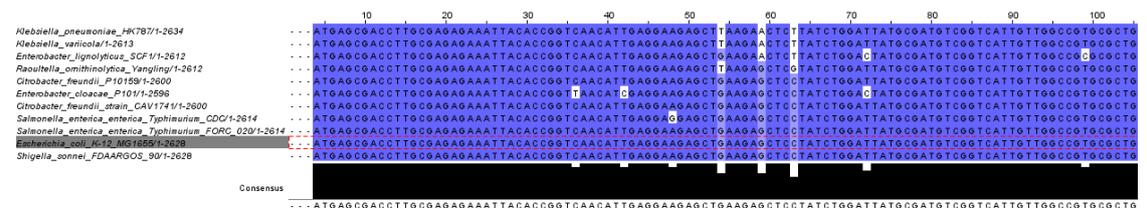


Figura 7 Visualização de uma região conservada do alinhamento de seqüências para o gene *gyrA*. A cor azul escuro representa a porcentagem de semelhança. Marcado em cinza, está a seqüência do gene *gyrA* de *Escherichia coli* utilizada como molde para o alinhamento.



Os genes *dnaG*, *gyrB*, *gyrA* e *secA* apresentaram resultados dentro dos parâmetros ideais de PCR, com no máximo três *mismatches* não estando nenhum deles até 3 posições próximas da terminação 3' dos *primers*, uma vez que isso seria prejudicial a reações de PCR. Já *era*, que codifica uma proteína

com atividade de *GTPase*, não foi capaz de produzir *primers* universais devido aos resultados revelando grande número de ampliações inespecíficas e *mismatches*.

Figura 8 Visualização parcial do alinhamento de sequência para o gene *dnaG* via *Jalview*.

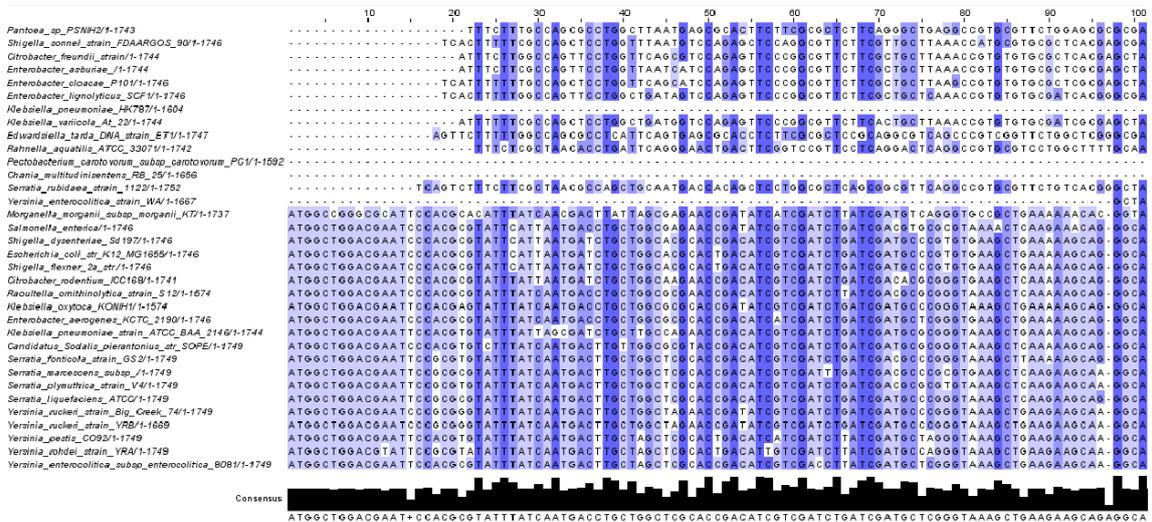
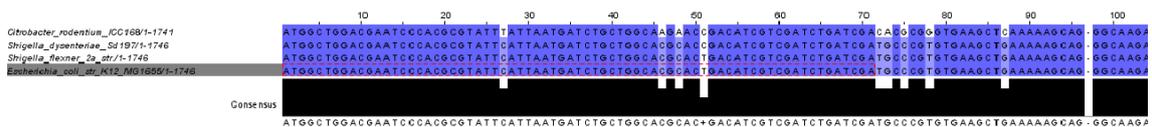
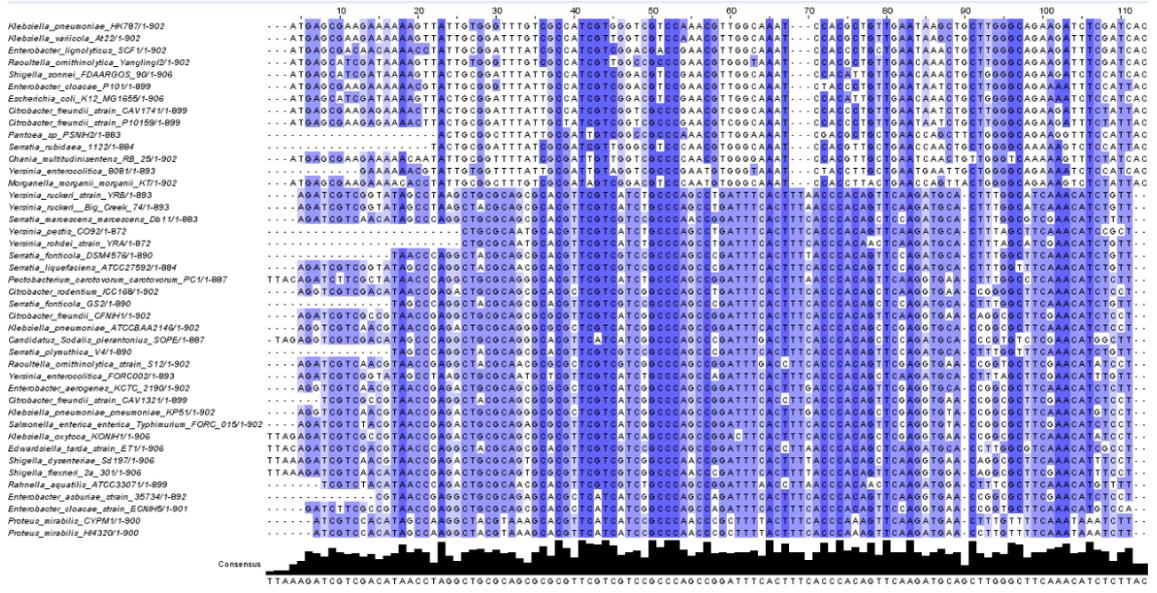


Figura 9 Representação de organismos que partilham maior semelhança para o gene *dnaG*, segundo *ClutalOmega*, via *Jalview*.



**Figura 10 Visualização parcial do alinhamento de seqüências para o gene *era* via Jalview.** Os tons azul, do mais claro ao mais escuro, representam o acréscimo da porcentagem de identidade. Predominam mismatches e gaps que inviabilizam regiões para construção de primers universais.



**Figura 11 Representação de organismos que partilham maior semelhança para o gene *era*, segundo ClustalOmega, via Jalview.**

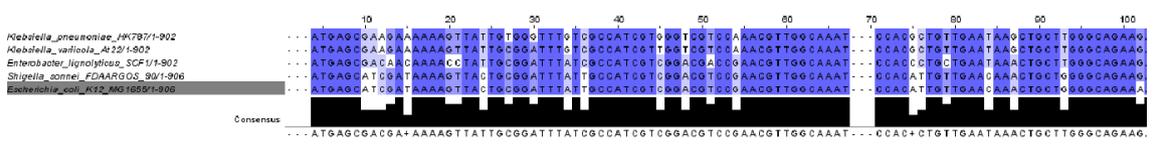


Figura 12 Visualização parcial do alinhamento de seqüências para o gene *secA* via Jalview.

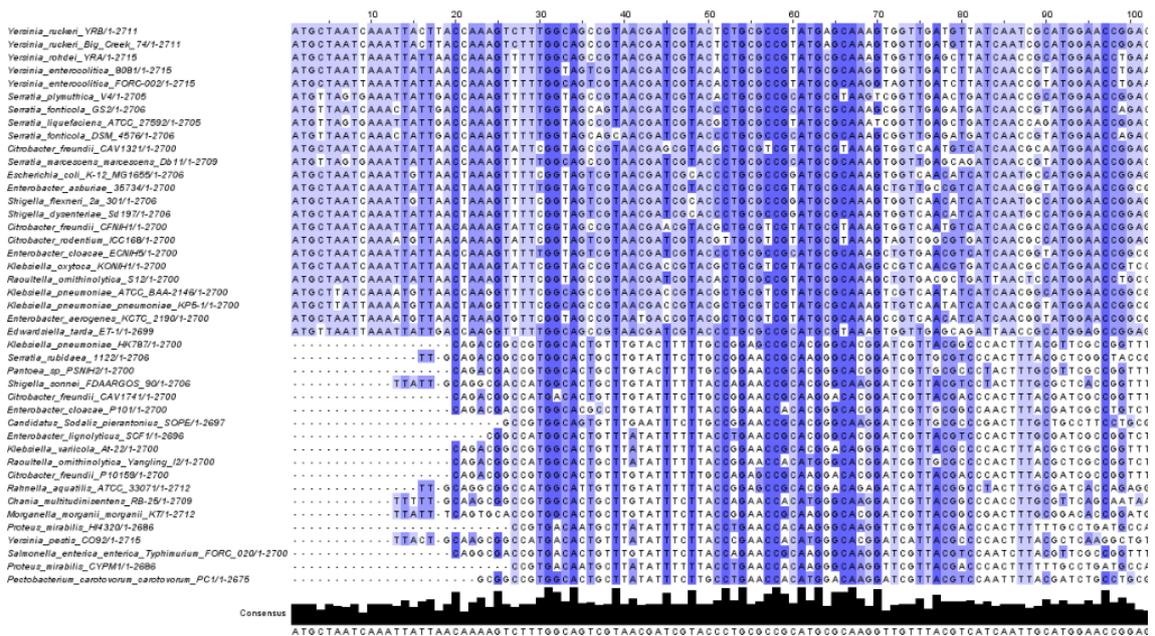
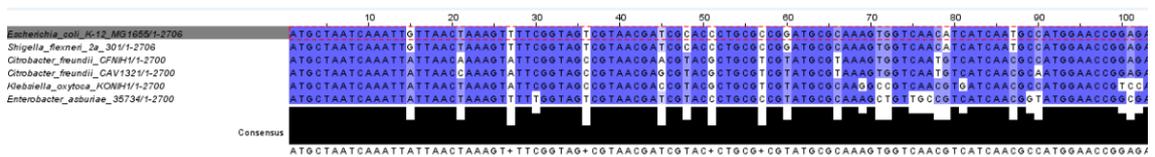


Figura 13 Representação de organismos que partilham maior semelhança para o gene *era*, segundo ClustalOmega, via Jalview.



Como não foi possível desenhar *primers* universais para *dnaG* e *era*, por conta de sua alta variabilidade e como foi observado para os genes *gyrA* e *gyrB*, foi estabelecido o uso dos 150 pares iniciais dos genes, sendo submetidos diretamente ao *PrimerQuest Tool*. O número de pares de *primers* possíveis variou entre 3 e 5 opções (Tabela 2). Todas as opções criadas, para todos os genes, foram testadas no *Primer-Blast* no filo *Proteobacteria*. O par de *primers*, de cada gene, que amplificasse para organismos de interesse clínico com um número e posição tolerável de *mismatches*, era mantido. Os demais foram desconsiderados. O gene *dnaG*, foi o único cujo todos resultados produzidos eram ruins o suficiente para serem totalmente excluídos (Tabela 3). Devido a prevalência de resultados dentro da família de enterobactérias, assumiu-se então a denominação “*Primers* de Enterobactérias”(Tabelas 2 e 3).

**Tabela 2 Primers de enterobactérias desenhados a partir dos genes *dnaG*, *gyrA*, *gyrB* e *secA* de *Escherichia coli* e suas propriedades técnicas.** Foram escolhidos entre os propostos pela ferramenta *PrimerQuest Tool*, após serem testados contra o banco de enterobactérias do NCBI pela ferramenta *Primer-Blast*. Todas as opções estão dentro dos parâmetros ideais comumente praticados dentro da metodologia de qPCR. Os *primers* para 16S foiram incluídos por serem comumente utilizados na literatura.

Sequência 5' – 3'	Início	Fim	Tamanho bp	Tm °C	GC%	Amplicon bp
<b><i>E. coli dnaG</i></b>						
Forward <u>GACGAATCCCACGCGTATT</u>	8	27	19	62	52.6	103
Reverse <u>TGGAAATTCTTGCCCTGCT</u>	92	111	19	62	47.4	
<b><i>E. coli gyrA</i></b>						
Forward <u>CGAGAGAAATTACACCGGTCA</u>	14	35	21	62	47.6	75
Reverse <u>CAATGACCGACATCGCATAATC</u>	67	89	22	62	45.5	
<b><i>E.coli gyrB</i></b>						
Forward <u>GTCCTGAAAGGGCTGGATG</u>	34	53	19	62	57.9	120
Reverse <u>CGAATACCATGTGGTGCAGA</u>	105	125	20	62	50	
<b><i>E. coli secA</i></b>						
Forward <u>GGTAGTCGTAACGATCGCA</u>	31	50	19	61	52.6	79
Reverse <u>TTCCATCTCCGGTTCCAT</u>	91	110	19	61	47.4	
<b><i>E. coli 16S</i></b>						
Forward <u>TCAAGTCATCATGGCCCTTAC</u>	1194	1215	21	62	47.6	111
Reverse <u>CGGACTACGACGCACTTTAT</u>	1285	1305	20	62	50	

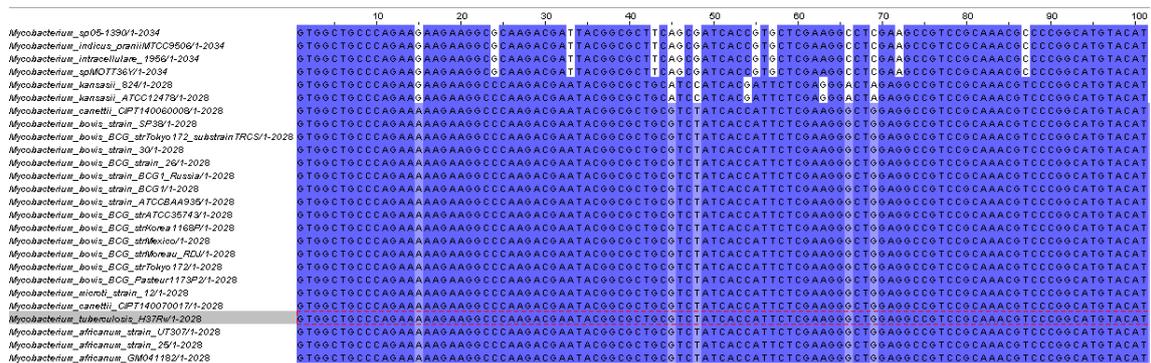
**Tabela 3 Tabela de amplificação baseada em PCR virtual com a ferramenta *Primer-Blast* de enterobactérias de interesse clínico.** Em verde está representado a confirmação que os iniciadores propostos amplificam o gene localizados nas colunas. Em vermelho, a sinalização de que o par de iniciadores não é capaz de amplificar o gene, ou não atendia às exigências de iniciadores ideais, comprometendo a reação de PCR. Toda amplificação com mais de 3 *mismatches*, ou com mismatches até 3 posições próximas da terminação 3' dos *primers* foram excluídas da análise. Assim como espécies que amplificavam de modo inespecífico.

<b>Espécies</b>	<b>gyrA</b>	<b>gyrB</b>	<b>dnaG</b>	<b>secA</b>
<i>Citrobacter amalonaticus</i>	Verde	Verde	Vermelho	Vermelho
<i>Citrobacter freundii</i>	Verde	Verde	Vermelho	Vermelho
<i>Citrobacter koseri</i>	Verde	Verde	Vermelho	Vermelho
<i>Citrobacter rodentium</i>	Verde	Verde	Verde	Vermelho
<i>Enterobacter aerogenes</i>	Verde	Verde	Vermelho	Vermelho
<i>Enterobacter agglomerans</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Enterobacter asburiae</i>	Vermelho	Verde	Vermelho	Vermelho
<i>Enterobacter cloacae</i>	Vermelho	Verde	Vermelho	Vermelho
<i>Enterobacter hormachei</i>	Vermelho	Verde	Vermelho	Vermelho
<i>Enterobacter lignolyticus</i>	Vermelho	Verde	Vermelho	Vermelho
<i>Enterobacter sp.</i>	Vermelho	Verde	Vermelho	Vermelho
<b><i>Escherichia coli</i></b>	Verde	Verde	Verde	Verde
<i>Klebsiella michiganensis</i>	Verde	Verde	Vermelho	Vermelho
<i>Klebsiella oxytoca</i>	Verde	Verde	Vermelho	Vermelho
<i>Klebsiella pneumoniae</i>	Verde	Verde	Verde	Vermelho
<i>Klebsiella quasipneumoniae</i>	Verde	Verde	Verde	Vermelho
<i>Klebsiella sp.</i>	Vermelho	Verde	Vermelho	Vermelho
<i>Klebsiella variicola</i>	Verde	Verde	Vermelho	Vermelho
<i>Morganella morganni</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Proteus mirabilis</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Proteus vulgaris</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Providencia spp.</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Salmonella typhi</i>	Verde	Verde	Vermelho	Vermelho
<i>Salmonella spp.</i>	Verde	Verde	Vermelho	Vermelho
<i>Serratia spp.</i>	Vermelho	Vermelho	Vermelho	Vermelho
<i>Shigella boydii</i>	Verde	Verde	Verde	Verde
<i>Shigella dysenteriae</i>	Verde	Verde	Verde	Verde
<i>Shigella flexeneri</i>	Vermelho	Verde	Vermelho	Verde
<i>Shigella sonnei</i>	Verde	Verde	Vermelho	Verde
<i>Shigella spp.</i>	Vermelho	Verde	Vermelho	Verde
<i>Yersinia enterocolitica</i>	Vermelho	Vermelho	Vermelho	Vermelho

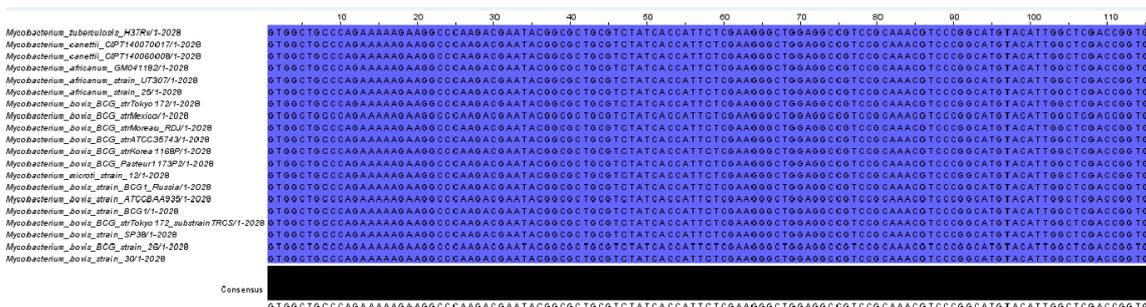
### 5.1.2 Primers para Complexo *Mycobacterium tuberculosis* (MTC)

No que se trata dos genes de *Mycobacterium tuberculosis*, quando observado o alinhamento através do Jalview, um padrão de divisão entre dois grupos foi observado (Figura 14). A homologia das sequências foi ainda mais acentuada do que o alinhamento entre proteobactérias, principalmente entre grupos de organismos classificados como complexo *Mycobacterium tuberculosis*. Os outros organismos, *Mycobacterium sp*, *Mycobacterium indicus*, *Mycobacterium kansasii*, *Mycobacterium intracellulare*, continuam variações o suficiente para desconsiderar qualquer abordagem de se desenhar *primers*. Constatou-se então que para espécimes MTC, era possível a aplicação da primeira estratégia. MTC engloba tradicionalmente *M. tuberculosis*, *M. africanum*, *M. canetti*, *M. bovis*, *M. microti*, *M. caprae*, *M. orygis* e *M. pinnipedii* (Figura 14). O MTC é caracterizado pelo incomum nível de conservação de seus genes *housekeeping* (de metabolismo basal) e sequências quase idênticas de RNA ribossômico 16S (BROSCH et al., 2002; HUARD et al., 2006; VAN INGEN et al., 2012; apud CHIMARA; FERRAZOLI; LEÃO, 2004; COSCOLLA et al., 2013).

**Figura 14** Visualização Parcial do alinhamento do gene *gyrB* dentro da família *Mycobacteriaceae* via Jalview.



**Figura 15** Visualização de uma região conservada do alinhamento de sequências para o gene *gyrB* na família *Mycobacteriaceae*.



**Tabela 4** Tabela de amplificação de espécies do Complexo *Mycobacterium tuberculosis* baseada em PCR virtual com a ferramenta *Primer-Blast*. Em verde está representado a confirmação que os iniciadores propostos amplificam o gene localizados nas colunas. Todos pares de *primers* propostos para todos os genes foram capazes de amplificar para todas as espécies do complexo. Toda amplificação com mais de 3 mismatches, ou com mismatches até 3 posições próximas da terminação 3' dos *primers* foram excluídas da análise. Assim como espécies que amplificavam de modo inespecífico.

Espécies	<i>gyrA</i>	<i>gyrB</i>	<i>dnaG</i>	<i>secA</i>	<i>era</i>
<i>Mycobacterium bovis</i>					
<i>Mycobacterium canettii</i>					
<i>Mycobacterium tuberculosis</i>					
<i>Mycobacterium africanum</i>					
<i>Mycobacterium microti</i>					

Para a segunda estratégia, utilizou-se também a sequência dos 150 pares de bases iniciais de cada gene para o desenho no *PrimerQuest Tool*. A quantidade de opções variou entre 4 e 6 opções, mas tendo em vista a alta homogeneidade entre os espécimes, assumiu-se a primeira opção para seguir com a análise. Quando testados pelo *Primer-Blast*, não importou contra qual banco de dados fosse aplicado, seja *Actinobacteria*, ordem *Corynebacteriales*, ou família *Mycobacteriaceae*, predominavam resultados MTC com 100% de identidade, ou seja nenhum *mismatch*. Portanto, as duas estratégias foram possíveis devido as características de semelhança entre os membros do MTC.

**Tabela 5 Primers para complexo *Mycobacterium tuberculosis* (MTC) desenhados a partir dos genes *dnaG*, *gyrA*, *gyrB* e *secA* de *Mycobacterium tuberculosis* e suas propriedades técnicas.** Foram escolhidos entre os propostos pela ferramenta *PrimerQuest Tool*, após serem testados contra o banco de espécies MTC do NCBI pela ferramenta *Primer-Blast*. Todas as opções estão dentro dos parâmetros ideais comumente praticados dentro da metodologia de qPCR.

Sequência 5' - 3'	Começo	Término	Tamanho bp	T <sub>m</sub> °C	GC%	Amplicon bp
<b><i>M. tuberculosis dnaG</i></b>						
Forward <u>TGTCGTCGCGACTATGT</u>	63	81	18	63	55.6	82
Reverse <u>GGACTTCTCGTTGTGAAACGG</u>	124	145	21	63	52.4	
<b><i>M. tuberculosis era</i></b>						
Forward <u>CGAATTCATTCTGGCTTTGTG</u>	6	28	22	62	45.5	101
Reverse <u>TCGACGTGATTGCCACTTT</u>	88	107	19	62	47.4	
<b><i>M. tuberculosis gyrA</i></b>						
Forward <u>TGACAGACACGACGTTGC</u>	2	20	18	62	55.6	107
Reverse <u>GATCAGCTCATCGCATAGT</u>	89	109	20	62	50	
<b><i>M. tuberculosis gyrB</i></b>						
Forward <u>AAGAAGGCCCAAGACGAATAC</u>	16	37	21	62	47.6	130
Reverse <u>CCTCCAAATGAGATGGTGTAA</u>	124	146	22	62	45.5	
<b><i>M. tuberculosis secA</i></b>						
Forward <u>TCAAGAAGGTGGCCGACTAT</u>	50	70	20	63	50	94
Reverse <u>CGCTTGAACCTCGTCGGTTT</u>	125	144	19	63	52.6	

### 5.1.3 Primers de organismos modelo em *Firmicutes* e *Proteobactéria*

Tabela 6 Tabela de primers dos genes *dnaG*, *secA*, *gyrB* e *16S* para diferentes organismos modelo em *Firmicutes* e *Proteobacteria*.

Espécie	Gene	Sentido	Sequência	Length	GC%	Tm °C
<i>B. subtilis</i>	dnaG	Foward	GGGAAATCGGATACCAGATGAA	22	45	58
		Reverse	GTTTCGGCCTTGCTTCTTTAAT	22	40	58
	gyrB	Foward	GTTGAAGTGGCTTTGCAATACA	22	41	58
		Reverse	CTTCATGGGTACCGCCTTC	19	58	59
	secA	Foward	ACCATATCAACCAGGCCTTAAA	22	40	57
		Reverse	CCCGTGAAGGAATCAACAATAAC	23	43	58
	16S	Foward	GTCGCAAGACTGAACTCAAAG	22	45	58
		Reverse	GAGGATGTCAAGACCTGGTAAG	22	50	58
<i>P. aeruginosa</i>	dnaG	Foward	GACAAGCCGAAGTACCTGAA	20	50	58
		Reverse	CCTCGTAGAGTCCGTAGAGTT	21	52	58
	gyrB	Foward	GTACGACTCTTCCAGCATCAA	21	48	59
		Reverse	CGCTTCGTGATGGAGTTAT	20	50	58
	secA	Foward	CCGAGCTGTACATCAAGATCAA	22	45	58
		Reverse	CTTCTCGTCGATGCTGTAGTG	21	52	58
	16S	Foward	CGCAACCCCTGTCTTAGTTA	21	47	58
		Reverse	GTAAGGGCCATGATGACTTGA	21	47	58
<i>S. enterica</i>	dnaG	Foward	GATACGCTGGTGCCTAAAGA	20	50	58
		Reverse	GCGGCAGGAGGCTATTTAAA	19	52	58
	secA	Foward	AAAGCGGGCATCAAACATAAC	21	42	58
		Reverse	GCCATGTTGGTAGCGATAGT	20	50	58
	16S	Foward	CCTCATAAAGTGCCTCGTAGTC	22	50	58
		Reverse	CGTATTCACCGTGGCATTCT	20	50	58
<i>S. aerus</i>	dnaG	Foward	GAAGATGAGTATGGCGGTTACA	22	45	58
		Reverse	AATGCTCGCTCCGCTTT	17	52	58
	gyrB	Foward	GCTGGGCAAATACAAGTATTAGAAG	25	40	62
		Reverse	CCCACACTAAATGGTGCAAAC	21	48	59
	secA	Foward	CGAAGGAAGGCGTTCAAATTC	21	47	58
		Reverse	CTGTCATACCCGCAAGTTTATTG	23	43	58
	16S	Foward	GCGACTTTCTGGTCTGTAAC	21	47	58
		Reverse	CCCTAACACTTAGCACTCATCG	22	50	58
<i>X. campestris</i>	dnaG	Foward	TACCTCAATTCGCCGAAAC	20	50	58
		Reverse	CGATCAGCCGTTTCGATCTTC	20	55	59
	gyrB	Foward	CAGGTGGGCACGCTGAT	17	60	64
		Reverse	GATGATGCGGTGGTAGCG	18	58	62
	secA	Foward	CAGAACCTGTTCCGCATGTA	20	50	58
		Reverse	GTAGATGCTCTGGAATTCGTAGG	23	47	58
	16S	Foward	AAGGCCTTCGGTTGTAAAG	20	50	58
		Reverse	CGAAGTTAGCCGGTGCTTAT	20	50	58

### 5.1.4 Análise *in silico* da estabilidade dos genes candidatos

Figura 16 Boxplot dos dados de *gyrB*, *gyrA*, *dnaG* e *secA* para *Escherichia coli*. Os dados foram obtidos na plataforma PATRIC e plotados usando o software GraphPad Prism 6. Submetidos a teste T de *student*, os genes *gyrB*, *16S*, *dnaG* e *secA* não apresentaram diferença estatística ( $P > 0,05$ ). Por sua vez, *era*, mostrou valor de  $p < 0,05$  entre as fases exponencial e estacionária.

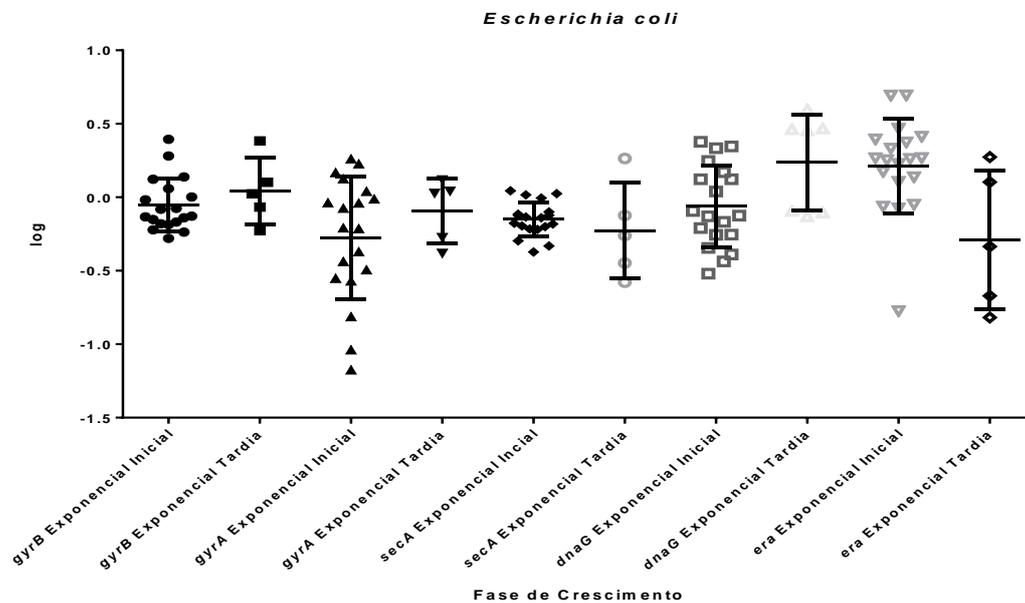
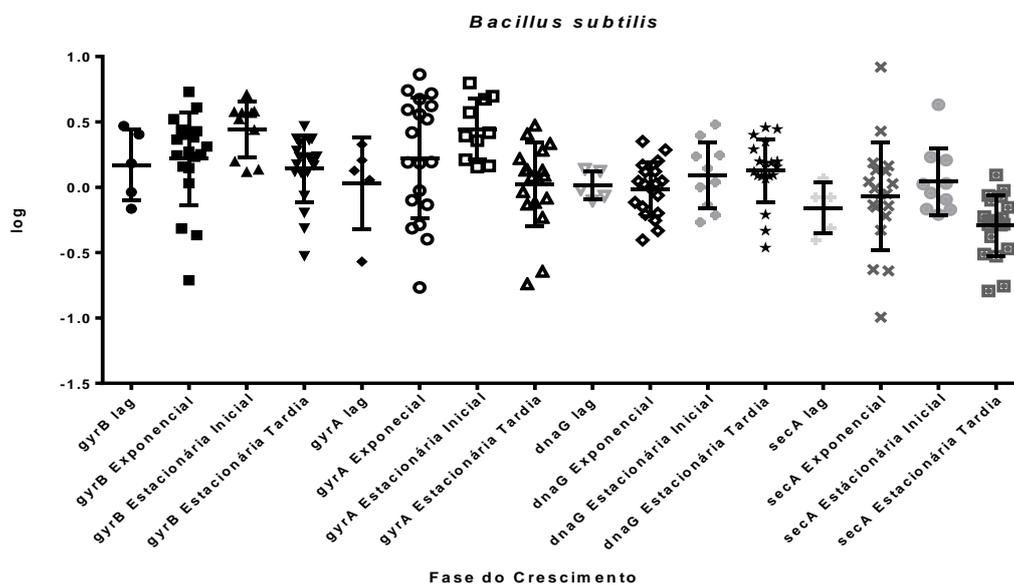


Figura 17 Boxplot dos dados de *gyrB*, *gyrA*, *dnaG* e *secA* para *Bacillus subtilis*. Os dados foram obtidos na plataforma PATRIC e plotados usando o software GraphPad Prism 6. Submetidos a teste T de *student*, os valores de expressão de *gyrB* não apontaram diferença estatística ( $P > 0,05$ ). Para *gyrA*, o valor de  $P < 0,05$  mostrou diferença estatística entre as fases estacionária inicial/tardia, e lag/estacionária. *dnaG* não se mostrou estável ( $P < 0,05$ ) somente para a comparação entre fase exponencial/estacionária. *secA* não mostrou estabilidade entre as fases estacionária inicial/tardia. Já *era* mostrou diferença de estabilidade entre as condições exponencial/estacionária.



### 5.1.5 Padronização da extração de RNA

A cultura de *E. coli* foi acompanhada até que atingisse em primeiro, a fase logarítmica de crescimento. Após 1h:30min a densidade ótica atingiu para as duplicatas RNA1 e RNA2 o valor de 0,57 e 0,64 respectivamente, e RNAH 0,63. Após 5h:30min de cultivo, os valores de densidade ótica para RNA1 e RNA2 eram de 2,58 e 2,65, respectivamente, e para RNAH, 2,63.

Após a extração de RNA as amostras foram dosadas por meio da *μDrop Plate*<sup>TM</sup> da *Thermo Scientific*<sup>®</sup> e dosadas mais uma vez depois da repurificação com *DNase* (Tabela 7). Houve um decréscimo da concentração de RNA, o que pode ser atribuído a perdas naturais durante a execução da técnica, ação de *RNases*, ou a digestão de *gDNA* pela ação da *DNase*. Os valores da relação A260/280 se mantiveram dentro do esperado para pureza de RNA. Os valores de A/260/230, por sua vez, variaram drasticamente (Tabela 7), talvez por resíduos de proteína ou de reagentes utilizados no processo de cleanup.

Tabela 7 Dados de quantificação e pureza do RNA obtido.

Amostra	Antes do Tratamento com DNase			Depois do Tratamento com DNase		
	CC ng/μL	A 260/280	A 260/230	CC ng/μL	A 260/280	A 260/230
RNA1 1:30h	41,65	2,29	1,74	36,1	2,29	0,17
RNA2 1:30h	36,7	1,52	1,94	16,12	2,23	0,28
RNA Heat Shock 1:30h	38,4	2,29	4,37	23,15	2,22	0,89
RNA1 5:30h	85,42	2,29	3,38	45,55	2,17	1,16
RNA2 5:30h	98,75	2,28	0,61	44,8	2,28	1,39
RNA Heat Shock 5:30h	97,27	2,17	1,32	47,07	2,24	1,23

## 5.2 RT-qPCR

### 5.2.1 Determinação da Eficiência dos *Primers*

A eficiência dos *primers* de cada gene foi determinada como descrito no item 4.5.3 pela construção de uma curva de diluição de 5 pontos (1x, 10x, 100x, 1000x e 10000x) (Figura 18). As médias dos valores de Ct de cada gene para cDNA1 e cDNA2 foram usadas para a construção de uma reta de regressão linear. Os valores de Cq que estavam acima dos que haviam detectado amplificação de DNA genômico no item 4.4.4, foram desconsiderados, gerando resultados mais acurados.

Os valores de eficiência para o gene *gyrB* foi 92,26%. Para *16S*, *dnaG* e *secA*, os valores foram de 93,07%, 100,29% e 103,28% (Tabela 9).

Um par de *primers* para *gyrB* desenhados diretamente na ferramenta online *PrimerQuest Tool*, sem considerar a janela de 150 pares de base estabelecida, também foi testado e, a critério de identificação, denominado *gyrBV* (Figura 18 E). Seu resultado corrobora com os do *primer* desenhado dentro dos critérios deste estudo, visto que seus resultados (Tabela 8) são significativamente próximos.

Figura 18 Curvas de diluição obtidas por qPCR para amostras diluídas 1x, 10x, 100x, 1000x e 10000x.

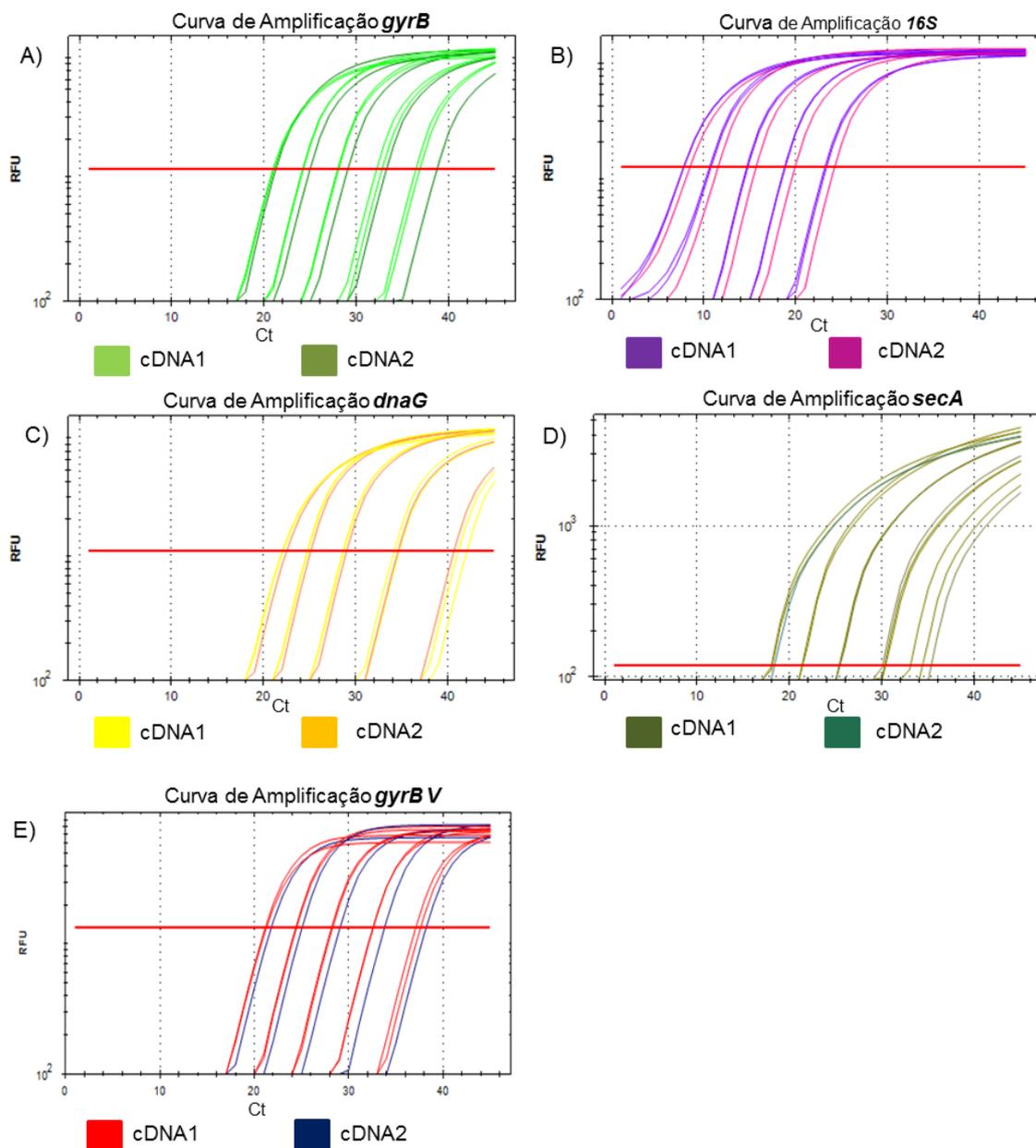


Tabela 8 Amostras e os valores de Ct obtidos por qPCR para cada diluição de cada par de *primer*.

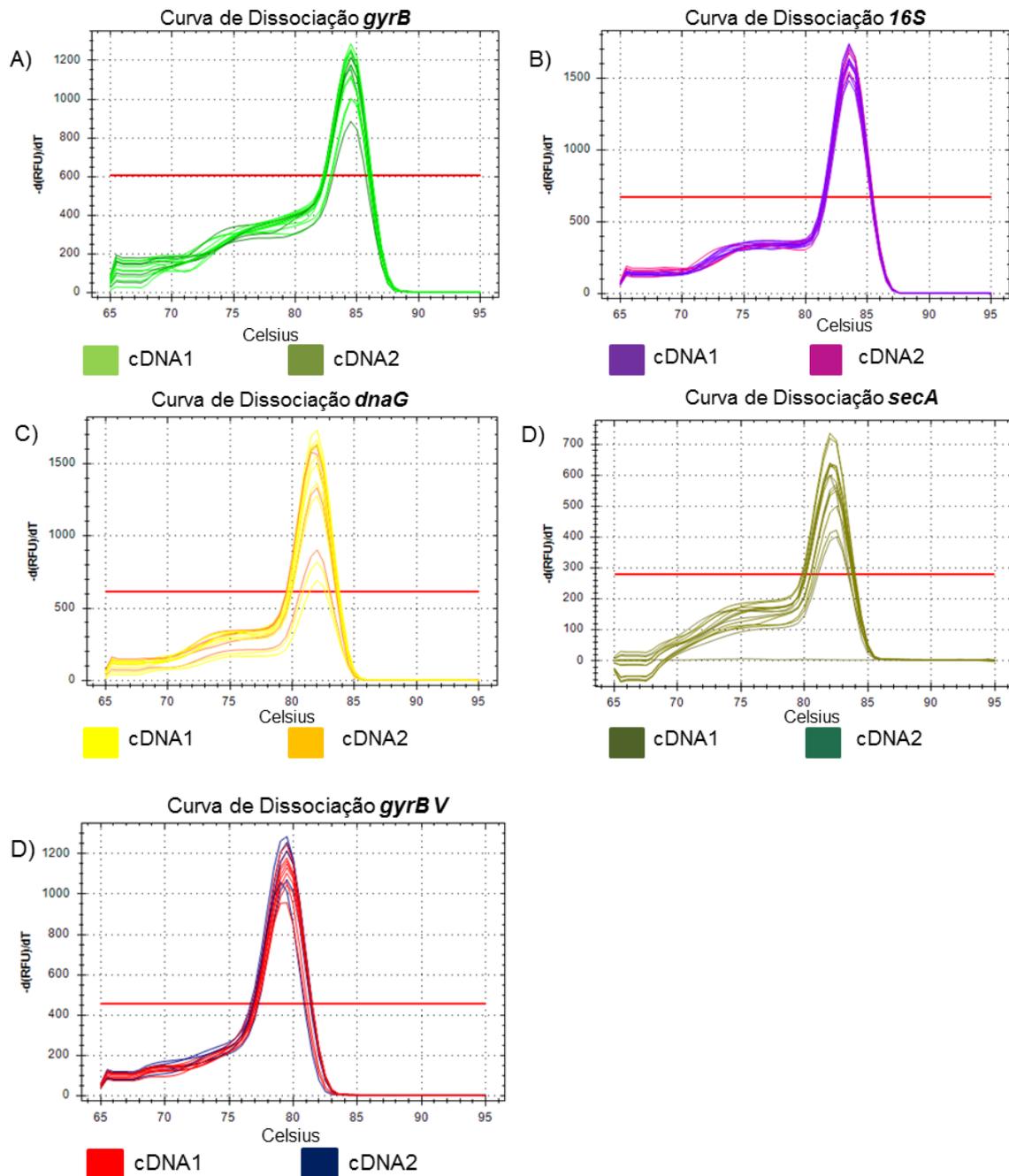
Gene	Diluição	cDNA1Ct	cDNA2	Média Ct
<i>gyrB</i>	100x	27,66	27,77	28,07
	10x	23,89	23,97	24,18
	1x	20,77	20,64	20,81
<i>gyrBV</i>	100x	27,72	27,59	27,94
	10x	23,94	23,83	24,09
	1x	20,55	20,67	20,81
<i>16S</i>	100x	14,25	14,34	14,62
	10x	10,16	10,3	10,53
	1x	7,28	7,31	7,54
<i>dnaG</i>	100x	28,38	28,44	28,54
	10x	24,56	24,5	24,67
	1x	21,83	21,73	21,94
<i>secA</i>	100x	30,83	30,87	30,89
	10x	26,79	26,41	NA
	1x	24,66	24,09	24,52

Tabela 9 Eficiência dos *primers* para os genes candidatos a referência e a inclinação das retas obtidas a partir dos valores de Ct.

Gene	Slope	R <sup>2</sup>	Eficiência
<i>gyrB</i>	-3,5225	0,99	1.92 (92,26%)
<i>gyrBV</i>	-3,505	0,99	1.93 (92,89%)
<i>16S</i>	-3,5	0,99	1.93 (93,07%)
<i>dnaG</i>	-3,315	0,99	2.0 (100,29%)
<i>secA</i>	-3,2375	0,96	2.04 (103,65%)

## 5.2.2 Especificidade das reações de qPCR através de curvas de dissociação

Figura 19 Curvas de dissociação para os genes *gyrB*, *16S*, *dnaG* e *secA*. No item D) encontram-se os dados de dissociação de um par de *primers* de *gyrB* desenhado fora dos critérios de estudo, que foi usado para comparar dados.



### 5.2.3 Análise de estabilidade dos diferentes genes

Os valores de  $C_q$  para os genes candidatos a referência dos 5 pares de *primers* avaliados neste estudo, foram convertidos em quantidades relativas, corrigidos pelos valores de eficiência de amplificação, como descrito no item 4.5.5. Esses valores normalizados foram importados para o *software NormFinder* para determinar a estabilidade de expressão dos genes candidatos a genes de referência. Quanto menor o valor, mais estável é a expressão do gene. O *software* sugeriu *dnaG* como o gene mais estável com valor quase próximo a zero (0,014), seguido por *secA* com o mesmo valor (0,014), *gyrB* (0,128), *gyrBV* (0,281), *16S* (0,351) (Figura 21). Como todos os genes nesse estudo já eram tidos como bons candidatos a referência e usados com certa frequência na literatura, todos eles apresentaram valores baixos de índice de estabilidade de expressão. O gene *16S*, no entanto, apesar de comumente utilizado como gene de referência em bactéria, vem sendo descrito como não confiável e frequentemente invalidado, devido a grande quantidade de transcrito quando comparado a um gene alvo (CARVALHO et al., 2014; LIU et al., 2013b; TURRONI et al., 2011; VANDESOMPELE et al., 2002).

**Figura 20** Boxplot dos dados de  $C_q$  entre as amostras das condições normal e as obtidas após estresse térmico de 47°C por 10min.

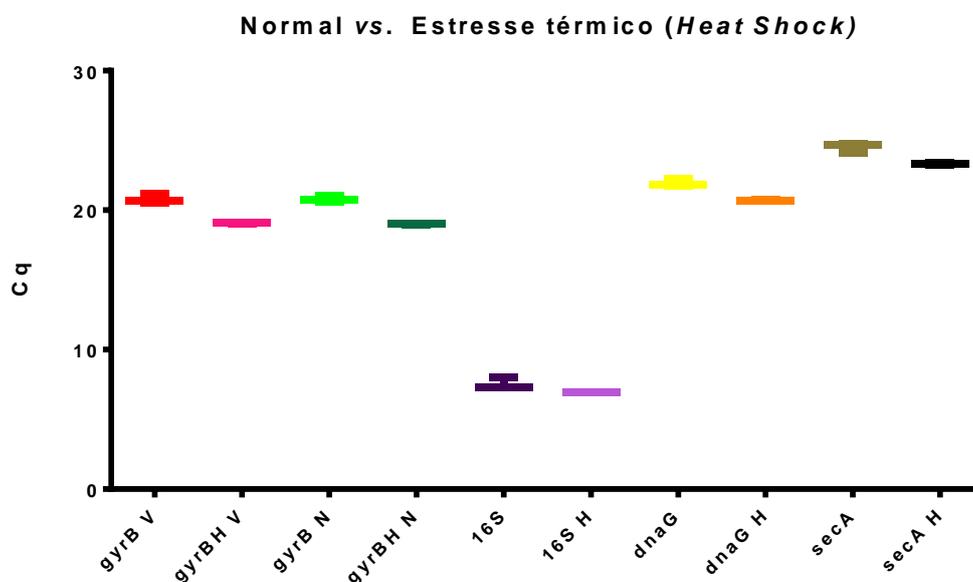


Figura 21 Estabilidade de expressão para os genes candidatos a gene normalizador utilizando o software *NormFinder*.

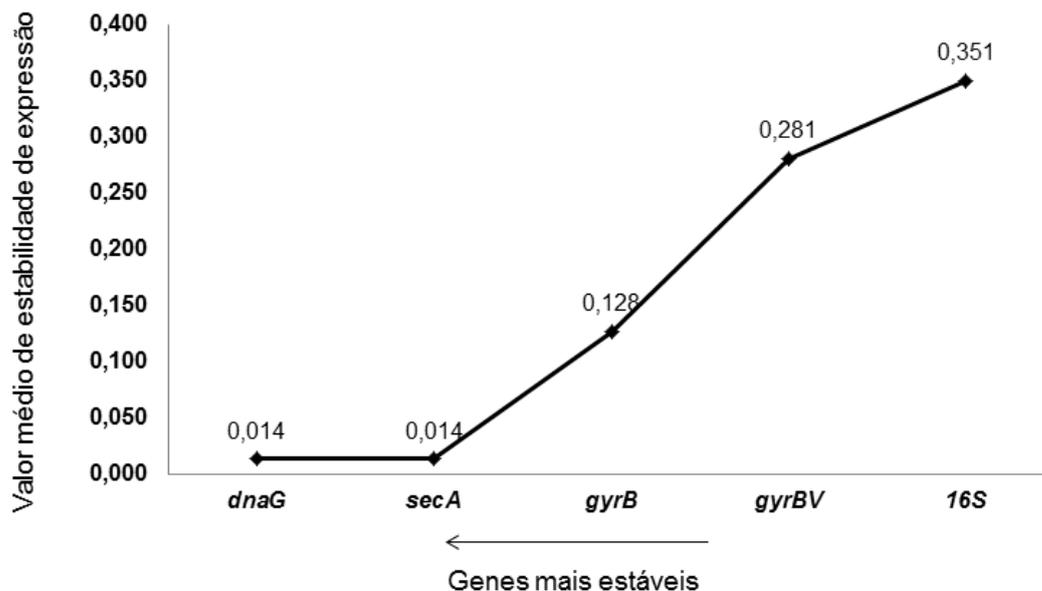


Tabela 10 Ranqueamento dos melhores genes de referência para os quatro principais métodos de acordo com a ferramenta RefFinder. O RefFinder faz uso dos dados brutos de Cq e portanto não leva em consideração a influência da eficiência dos *primers* utilizados no ensaio de qPCR. Assume, portanto que a eficiência de todos os *primers* seria de 100%.

Ordem de <i>Ranking</i> (Melhor--Bom--Mediano)					
Método	1	2	3	4	5
<b>Delta CT</b>	<i>dnaG</i>	16S	<i>gyrBV</i>	<i>secA</i>	<i>gyrBN</i>
<b>BestKeeper</b>	<i>secA</i>	<i>gyrBV</i>	<i>dnaG</i>	16S	<i>gyrBN</i>
<b>Normfinder</b>	16S	<i>dnaG</i>	<i>gyrBV</i>	<i>secA</i>	<i>gyrBN</i>
<b>Genorm</b>	16S/ <i>dnaG</i>		<i>gyrBV</i>	<i>secA</i>	<i>gyrBN</i>
<b>Ranking Recomendado</b>	<i>dnaG</i>	16S	<i>gyrBV</i>	<i>secA</i>	<i>gyrBN</i>

**Figura 22 Relação de estabilidade dos genes candidatos a referência entregue pela ferramenta online RefFinder.** Apesar de não levar em consideração a eficiência dos *primers* utilizados, os resultados corroboram com os apresentados pelo NormFinder, indicando *dnaG* como o gene mais estável entre os candidatos a referência.



Para a ferramenta online *RefFinder*, que faz uso dos dados brutos de Cq (XIE et al., 2012), sem considerar a eficiência dos *primers*, a comparação de múltiplas metodologias indicaram o gene *dnaG* como mais estável e mais recomendado (Figura 22), corroborando o resultado encontrado com *NormFinder* anteriormente. Individualmente, os gene mais estável para o método Delta-Cq foi *dnaG*, para o *BestKeeper*, *secA*, *NormFinder*, 16S, e a combinação 16S/*dnaG* segundo o algoritmo do *geNorm* (Tabela 10).

## 6. DISCUSSÃO

### 6.1 Desenho de *Primers*

#### 6.1.1 *Primers* Universais

Um par de *primers* universais deve ser capaz de amplificar um gene alvo em diferentes espécies de um gênero, família, ou grupos de hierarquias taxonômicas maiores. O número de espécies e agrupamento taxômico onde estes *primers* se aplicam, está diretamente relacionado a conservação do gene em questão. Para qPCR, um amplicon ideal deve ter entre 80 e 150 pares de bases (THORNTON; BASU, 2011). Sendo assim, é imprescindível a presença de ao menos uma região conservada de no mínimo 150/200 pares de base que permita o desenho de *primers* com o menor números de mismatches que não interfiram na eficiência e sensibilidade de um ensaio de qPCR.

A literatura também sugere que, em estudos de análise de expressão de diversos genes, o tamanho de todos os *amplicons* deve ser semelhante. Visto que, substâncias como SYBR<sup>®</sup>, vão produzir fluorescência de maior intensidade em produtos de amplificação maiores (THORNTON; BASU, 2011).

#### 6.1.2 *Primers* para *Enterobacteria*

O filo *Proteobacteria* inclui bactérias bem conhecidas, como a fermentativa *Escherichia coli*, e a aeróbica *Pseudomonas fluorescens*, a fotossintética *Chromatium* e muitas outras (KERSTERS et al., 2006). Comporta uma das maiores divisões dentro dos procariotos e acomoda a grande maioria das bactérias gram-negativas conhecidas (STACKEBRANDT; MURRAY; TRUPER, 1988 apud GUPTA, 2000). É também um grupo de grande importância clínica, contendo o maior número conhecido de patógenos humanos, animais e de plantas, contendo mais de 200 gêneros diferentes (HOLT et al., 1994 apud GUPTA, 2000).

Essa diversidade de organismos se reflete no nível de conservação dos genes *dnag*, *era*, *gyrA*, *gyrB*, *secA* e no alinhamento dessas sequências durante a aplicação da primeira estratégia de desenho de iniciadores. É por essa razão que nos resultados de alinhamento no *BLASTn* predominavam resultados para família *Enterobacteriaceae*. Devido ainda à grande quantidade

de sequências diferentes para cepas de *Escherichia coli* depositadas no NCBI, foi preciso excluir a espécie dentro da pesquisa.

A ferramenta *Jalview* é de grande utilidade para auxiliar a visualização de alinhamentos de sequência feitos em outras ferramentas. Durante as primeiras tentativas de alinhamento das sequências obtidas, verificou-se a necessidade de se fazer reajustes. Por vezes, foi preciso omitir algumas dessas sequências na visualização, ou ainda requerendo a completa exclusão de algumas espécies e, por consequência a realização de um novo alinhamento. Com isso, fragmentos ou sequências de menor similaridade tinham menos impacto no método e diminuía a inserção de *gaps* pelo *ClustalOmega*, no esforço de se obter o máximo de similaridade. Dessa forma foi possível encontrar regiões que atendessem de maneira satisfatória o tamanho necessário de nucleotídeos (entre 80 e 200 pares de base) e com o número de *mismatches* que não prejudicassem o desenho de *primers*. Assim que o número total de espécies possíveis para atender os requisitos de um *primer* universal foi se afunilando a depender da conservação gênica interespecie.

Dentre os genes estudados, o que apresentava maior conservação, e por consequência abrangia o maior número de espécies dentro dos parâmetros ideais para este estudo, foi o *gyrB* (Figura 5). A maior região conservada foi a porção inicial do gene, com cerca de 100 pares de base e é facilmente vista através do *Jalview*. Já o gene *gyrA*, demonstrou um padrão de conservação semelhante na região inicial do gene com cerca de 160 pares de base, como descrito no item 5.1 (Figura 7). Os genes *gyrA* e *gyrB* codificam ambas as subunidades da enzima DNA girase (topoisomerase II), essencial para a replicação, ou seja, pertencem ao metabolismo basal celular, atendendo uma das características de um gene candidato a referência (KOZERA; RAPACZ, 2013). São ainda conhecidos pelo alto grau de conservação entre espécies e podem ser utilizados na identificação de bactérias gram-negativas, em especial enterobactérias, via variações em porções do gene *gyrB* (BONASERA; ASSELIN; BEER, 2014; CHIMARA; FERRAZOLI; LEÃO, 2004).

A partir dos outros genes, a conservação gênica foi um fator crítico de limitação. O número de espécies que, através do alinhamento, exibiam regiões conservadas de tamanho suficiente para não comprometer a eficiência de

possíveis *primers*, foi reduzida de 6 para *secA* a nenhuma para *dnaG* e *era*. A estratégia até então utilizada, poderia estar sofrendo de alguma limitação que impedisse de se encontrar regiões de homologia interespecie.

Portanto, foi necessária a adoção de uma nova estratégia. Tendo visto os resultados anteriores, adotou-se o uso da região inicial, com 150 pares de bases, para todos os genes. Após o desenho dos *primers* (item 4.2), todas as sequências *primeras* foram submetidas ao *Primer-Blast* contra todo o banco do filo *Proteobacteria*. Dessa forma, era possível abranger o maior número de espécies possível que amplificasse a região de interesse e o par de iniciadores que tivesse o maior número de espécies, que não tivesse ampliações inespecíficas e mismatches em número tolerável, eram selecionados e os demais excluídos de análises futuras. Então, *era* foi o único gene cujo nenhum dos resultados atendia o critério necessário.

Para facilitar a identificação de espécies de interesse clínico ou biotecnológico, uma tabela foi criada contendo as principais enterobactérias. Assim, verde identifica que é possível a amplificação por meio do *primer* desenhado e em vermelho as espécies em que não é possível. A tabela também é outra forma de se visualizar a conservação dos genes candidatos a referência entre as espécies (Tabela 3).

É importante salientar que apenas três espécies amplificariam com excelência utilizando os iniciadores propostos. Contudo, isso não é necessariamente um problema, visto que até mesmo a utilização de um único gene de referência, num ensaio de análise de expressão, é aceitável quando apresentadas evidências claras de sua expressão estável dentro das condições estudadas. O número ideal de genes de referência pode ainda variar com a abordagem algorítmica utilizada para sua validação (ANDERSEN; JENSEN; ORNTOFT, 2004; VANDESOMPELE et al., 2002).

### **6.1.3 Iniciadores para complexo *Mycobacterium tuberculosis* (MTC)**

Tuberculose (TB) é causada por bactérias gram-positivas muito próximas, filogeneticamente conhecidas como *Mycobacterium tuberculosis complex* (MTC) (COLE et al., 1998). Apesar da proximidade genética, elas diferem em patogenicidade, distribuição geográfica, preferência de hospedeiro, e

importância para tuberculose humana (CHIMARA; FERRAZOLI; LEÃO, 2004). MTC inclui *M. tuberculosis*, *M. africanum*, *M. canetti*, *M. bovis*, *M. microti*, *M. caprae*, *M. orygis* e *M. pinnipedii* (BROSCH et al., 2002; COSCOLLA et al., 2013; HUARD et al., 2006; VAN INGEN et al., 2012).

Geneticamente, os membros do MTC compartilham uma conservação incomum de seus genes *housekeeping* e sequências quase idênticas de RNA ribossômico 16S (BROSCH et al., 2002; CHIMARA; FERRAZOLI; LEÃO, 2004). A proximidade filogenética das espécies foi visualizada em ambas as estratégias usadas para desenhar os *primers*. Quando submetidas ao *BLASTn*, as sequência dos genes *dnaG*, *era*, *gyrA*, *gyrB* e *secA*, sempre geravam resultados cuja a presença de MTC predominava, independente do nível hierárquico de taxonomia escolhido no banco de dados, filo *Actinobactéria*, ordem *Corynebacteriales*, ou família *Mycobacteriaceae*. No *Jalview* verificava-se a presença de grandes regiões conservadas para todos os genes candidatos a referência. Essa similaridade foi também representada nas imagens do alinhamento obtidas pelo *Jalview* e pelas características dos iniciadores (ausência de *mismatches* e 100% de identidade para MTC) quando testados pelo *Primer-Blast*. Portanto, as duas estratégias foram possíveis, principalmente pelas características de semelhança entre os membros do MTC, que, por sua vez, possuem grande importância clínica e interesse em pesquisa de expressão gênica. O interesse é caracterizado pela grande quantidade de publicações que procuram entender a patologia e fatores de virulência. Como para as enterobactérias, a alta conservação do gene *gyrB* pode ainda ser aplicada na diferenciação entre espécies do MTC via polimorfismo (CHIMARA; FERRAZOLI; LEÃO, 2004; KASAI; EZAKI; HARAYAMA, 2000).

#### **6.1.4 Primers de organismos modelo para *Firmicutes* e *Proteobacteria***

A mesma estratégia que gerou os *primers* para a família *Enterobacteriaceae*, foi aplicada utilizando como molde, diferentes organismos de importância clínica em *Proteobacteria* e *Firmicutes*. Por meio de PCR virtual, através da ferramenta *Primer-Blast*, os *primers* seriam capazes de amplificar para outras espécies, apesar de seu objetivo principal tenha sido o de se avaliar a estabilidade dos genes em diferentes organismos bacterianos.

## 6.2 Análise *in silico* da estabilidade dos genes candidatos

Antes de partir experimentalmente para os testes de estabilidade, decidiu-se por avaliar o comportamento dos candidatos a gene de referência *in silico*. No banco de dados online do *Pathosystems Resource Integration Center* estão depositados milhares de dados de transcriptoma e RT-qPCR sobre análise de expressão de vários organismos bacterianos e em diferentes condições experimentais (WATTAM et al., 2017).

Procuramos por experimentos “time point”, ou seja, que avaliassem os níveis de expressão gênica ao longo do crescimento bacteriano. Dentro desses experimentos, os separamos por fase de crescimento (exponencial inicial, exponencial tardia, estacionária inicial e estacionária tardia).

Os organismos avaliados foram: *Escherichia coli*, *Bacillus subtilis*, *Burkholderia pseudomallei*, *Helicobacter pylori*, *Pseudomonas aureuginosa*, *Salmonella enterica*, *Staphylococcus aureus* e *Streptococcus mutans*. Somente *E. coli* e *B. subtilis* possuíam dados com informação suficiente para seguir com a análise estatística. Ora por carência de dados ou ausência de identificação sobre o que tipo de experimento se tratava, as demais espécies foram desconsideradas.

Em *E. coli*, os genes, *gyrA*, *gyrB*, *secA* e *dnaG* não apresentaram diferenças estatísticas entre os dados da fase exponencial inicial e exponencial tardia, sugerindo mais uma vez a característica de referência para esses genes. Infelizmente, não haviam outros dados para outras fases do crescimento bacteriano. A exceção foi o gene *era*, cujo teste T indicou que houve diferença entre os grupos.

*B. subtilis*, por sua vez, possuía maior quantidade de dados, para fase *lag*, fase exponencial e fase estacionária de crescimento. Os testes T realizados entre os dados dessas condições revelaram que: para as comparações entre todas as fases de crescimento bacteriano, não houve diferença estatística para *gyrB*. Para *gyrA*, houveram diferença estatística entre as fases estacionária inicial/tardia, e lag/estacionária. *dnaG* não se mostrou estável somente para a comparação entre fase exponencial/estacionária. *secA* não mostrou estabilidade entre as fases estacionária inicial/tardia. Já *era*

mostrou diferença de estabilidade entre as condições exponencial/estacionária. O gene *era* então, devido ao seus resultados em ambas as espécies, foi excluído das futuras análises.

Em razão do bom desempenho de *gyrB*, *gyrA* foi desconsiderado de análises futuras por ambos codificarem para subunidades da proteína DNA gyrase e serem, portanto, co-regulados (KOZERA; RAPACZ, 2013). Essa co-regulação poderia influenciar análises estatísticas em qPCR e levar a interpretações equivocadas sobre suas características de gene de referência (ROBLEDO et al., 2014).

### **6.3 Estabilidade dos genes selecionados em resposta ao estresse térmico em *Escherichia coli***

Para esse desenho experimental, escolhemos o estresse térmico ou *Heat Shock* como condição de estresse devido a sua facilidade em ser empregada. Também é conhecido que, durante a exposição a altas temperaturas, ocorre a mobilização da transcrição de diversas proteínas *heat shock* que são responsáveis por proteger a célula principalmente contra danos estruturais (TOMOYASU; BUKAU; ARSENE, 2000). Num elavamento moderado de temperatura (30°C para 42°C), a síntese de algumas dessas proteínas é elevada entre 10 a 15 vezes nos primeiros 5 min de exposição e decresce após cerca de 25min (CHUNG; BANG; DRAKE, 2006). Sendo assim, foi decidido que a condição de estresse térmico teria no máximo 10° C de diferença da temperatura ideal de cultivo de 37°C e que 10min seria tempo o suficiente para que o calor se distribuísse igualmente pela cultura em tubo *Falcon* de 50mL.

O recrutamento da transcrição de todas essas proteínas por estresse térmico e as condições de ambientais da abundância de nutrientes na fase logarítmica e carência na fase estacionária deveriam ser ótimas para testar os genes candidatos a referência.

### **6.4 Extração de RNA e verificação de DNA genômico**

Uma determinação apurada da concentração de RNA é um fator crítico para a quantificação de mRNA em qPCR onde os número de cópias são normalizados contra RNA total. Qualquer contaminação significativa por gDNA resultará em imprecisão (BUSTIN, 2002). Por isso a padronização da

extração de RNA e a redução de gDNA provou-se o maior desafio deste estudo (Figura 3). Para alcançar resultados minimamente viáveis, requeriu-se a combinação de três diferentes kits comerciais, *TRIZO*<sup>®</sup>, *Rneasy*<sup>™</sup> mini kit da *QIAGEN*<sup>®</sup> e *SV Total RNA Isolation System*<sup>™</sup> da *Promega*<sup>®</sup>.

Quando empregados individualmente, os kits comerciais de coluna, não foram capazes de eliminar o DNA genômico com eficiência. Mesmo com o tratamento de *DNase* sobre a coluna de extração, testados, tanto no protocolo do fabricante, quanto em outros ajustes de tempo e temperatura. A contaminação por gDNA era consideravelmente maior na fase estacionária que na fase logarítmica do crescimento. Os manuais dos fabricantes não recomendavam os métodos para extração em amostras bacterianas em fase estacionária. Então, nenhum ajuste em volume de amostra ou número de células utilizado foi eficiente para reduzir a contaminação para níveis aceitáveis. Ao se introduzir a separação por *TRIZO*<sup>®</sup> antes da lise enzimática e passagem pela coluna de sílica, os resultados de pureza melhoraram significativamente. No entanto, o qPCR ainda demonstrava níveis de contaminação para as amostras da fase estacionária que inviabilizavam a validação do método. Num esforço de se utilizar estas amostras, foi incluída uma etapa de repurificação da amostra, utilizando o protocolo da *QIAGEN*<sup>®</sup> e a *DNase* inclusa no kit da *Promega*<sup>®</sup>. Foi então, que para as amostras de RNA da fase logarítmica que o gDNA foi reduzido até 1000x, dentro, portanto, de níveis aceitáveis.

## 6.5 RT-qPCR

Para as reações de qPCR nós escolhemos o kit comercial *GoTaq*<sup>®</sup> qPCR *Master Mix* da *Promega*<sup>®</sup> cujo sistema contém um corante fluorescente intercalante de DNA que emite mais fluorescência ao ligar-se a DNA dupla-fita que *SYBR*<sup>®</sup> *Green I*. É também simples de se usar e sua formulação inclui todos os componentes para qPCR, exceto molde de DNA e água.

### 6.5.1 Eficiência dos primers

Todos os *primers* testados apresentaram curva de amplificação como esperado, com excessão de *secA*, cuja a amplificação aconteceu menos como uma curva e mais como uma reta (Apêndice Figura 23). À primeira vista, isso

poderia indicar problemas com a temperatura do par de *primers* utilizado e os parâmetros do qPCR. No entanto, quando as curvas foram corrigidas levando em consideração a contaminação de gDNA o valor de eficiência de *secA* foi um pouco superior a 100% e os gráficos de dissociação demonstrava picos de temperatura muito semelhantes entre si (Tabela 9).

### 6.5.2 Estabilidade dos genes candidatos a referência

No gráfico gerado pela qPCR (Apêndice, Figura 23), as curvas de amplificação entre as amostras da fase logarítmica e àquela submetida a estresse térmico não apresentou grande diferença entre seus valores de Cq. O gene que apresentou a maior diferença entre as condições normais e de estresse (cerca de 2 ciclos) foi *gyrB*, para ambos os *primers* testados. Contudo, pequenas diferenças são esperadas devido a erros na técnica, tais como manipulação e degradação da amostra por RNase e diferenças sutis nos volumes entre pipetagens.

Quando os dados de Cq são apresentados em Boxplot (Figura 20) não se observa diferenças significativas entre as amostras obtidas na fase logarítmica em condições normais de crescimento para aquelas submetidas a estresse térmico. Contudo, pequenas diferenças são esperadas devido a erros na técnica, tais como manipulação e degradação da amostra por RNase e diferenças sutis nos volumes entre pipetagens.

A análise do algoritmo *NormFinder* escolheu o gene *dnaG*, que codifica a proteína DNA Primase, como o mais estável entre os candidatos, dentro do desenho experimental deste estudo. No entanto, é importante considerar que todos os genes tiveram pequenos valores de estabilidade (S). E como o *NormFinder* faz uso de dados linearizados, calculados de acordo com a eficiência dos *primers*, os genes *dnaG* e *secA* são reflexos de sua eficiência de amplificação de pouco mais de 100% (Figura 21).

Fazendo uso dos dados brutos de Cq, a ferramenta online *RefFinder* (XIE et al., 2011) comparou todos os algoritmos mais utilizados e indicou os genes *dnaG* e *16S* como os mais estáveis, corroborando com o resultado obtido exclusivamente com o *NormFinder*. Entretanto, a escolha do gene *16S* como um bom normalizador é reflexo da consideração do *RefFinder* de que todos os

*primers* possuem 100% de eficiência, o que não corresponde à realidade (Tabela 10; Figura 22). O gene *dnaG* não apresentou conservação de nucleotídeos o suficiente entre os organismos neste estudo, o que o impossibilita de ser caracterizado como gene universal. Mas seus dados de estabilidade corroboram com os de bancos de dado de transcriptoma bacteriano, onde, num conjunto de 1000 estudos, 92,5% validaram sua estabilidade em diferentes condições experimentais (ROCHA; SANTOS; PACHECO, 2015). Um estudo recente envolvendo *Herbaspirillum rubrisubalbicans* apontou através do software *NormFinder*, *dnaG* e *gyrA* como a melhor combinação de genes de referência para as condições avaliadas (POLESE et al., 2017).

O gene *16S*, como já explanado, é comumente usado como normalizador e foi decidido incluí-lo nas análises para avaliar sua estabilidade no desenho experimental do estudo. Porém, seus resultados satisfatórios podem não ser reais devido a problemas previamente descritos, tais como o número de cópias dentro do genoma e a grande quantidade de transcritos na célula quando comparado aos demais genes estudados (LI et al., 2015; VANDESOMPELE et al., 2002; XIE et al., 2012).

O gene *gyrB* é conhecido por apresentar estabilidade em até 95% dos 1009 experimentos bacterianos avaliados por Rocha et al., usando a plataforma *Pathosystems Resource Integration Center* (PATRIC) (ROCHA; SANTOS; PACHECO, 2015). Também é conhecido pela sua conservação entre enterobactérias e micobactérias (BONASERA; ASSELIN; BEER, 2014; CHIMARA; FERRAZOLI; LEÃO, 2004), o que faz deste gene um excelente candidato a referência em estudos de análise de expressão relativa em bactérias. Um estudo recente de validação de genes de referência em organismos como *Thermus sp.* demonstra a estabilidade de *gyrB* em condições aeróbicas e anaeróbicas fazendo uso dos software *NormFinder* e *BestKeeper* (CUSICK et al., 2015). Para *Avibacterium paragallinarum*, *gyrB* também configurava em boas posições de ranking, segundo diferentes softwares de normalização (WEN et al., 2016). Entretanto, outros estudos apontaram *gyrB* como gene instável (PESSOA et al., 2016; POLESE et al., 2017; VANDRISSE; PARKS; ESCALANTE-SEMERENA, 2017). Isto é uma demonstração de que

para cada organismo e condições experimentais utilizadas em experimentos de análise de expressão relativa, é necessária a validação apropriada de genes de referência (BUSTIN et al., 2009)

Ao contrário do que foi recomendado pelo MIQE *Guidelines* (BUSTIN et al, 2009), a maioria das publicações tem testado mais de um algoritmo para validar genes de referência, que algumas vezes apresentam um conjunto de genes diferentes para cada algoritmo (ROCHA; SANTOS; PACHECO, 2015). Mas, na maioria dos casos, estabelecendo genes que apresentam bom desempenho estatístico como opções para normalização entre as ferramentas. A escolha de mais de uma abordagem matemática pode estar relacionada ao conhecimento de que podem gerar resultados distintos e que, por comparação, poderia se encontrar o gene normalizador ideal entre todas as ferramentas.

## 7. CONSIDERAÇÕES FINAIS

O uso de genes de referência validados experimentalmente é um passo crítico para assegurar a veracidade dos dados de RT-qPCR. Faz-se necessária uma avaliação sofisticada e criteriosa da bibliografia utilizada como referência e o uso de guias para busca dos melhores genes de referência para o organismo e condição estudada.

A maior barreira para o desenvolvimento de um conjunto de *primers* universais bacterianos para RT-qPCR é a conservação dos genes trabalhados. Sem uma região de tamanho entre 80 e 200 pares de base com mínima quantidade de *mismatches* não é possível a construção de *primers* universais. Os genes *dnaG*, *gyrA*, *secA* e *era*, por sua vez, não apresentavam homologia suficiente entre as espécies dentro do filo *Proteobacteria* e família *Enterobacteriaceae* para permitir o desenho de *primers* universais dentro dos critérios do estudo. Entretanto, todos os genes apresentavam homologia suficiente entre as espécies do complexo *Mycobacterium tuberculosis*, cuja proximidade genética é conhecida.

O gene *gyrB* foi o gene mais conservado dentro das análises e pode ser usado como único normalizador, caso sua estabilidade seja validada experimentalmente (ANDERSEN; JENSEN; ØRNTOFT, 2004). Contudo, quando comparados os valores de expressão dentro das condições avaliadas, o gene mais estável foi o *dnaG*.

A comparação de dados de estabilidade entre diferentes algoritmos tem sido uma tendência entre as publicações após o MIQE guidelines. O desenvolvimento do *RefFinder* é um exemplo do esforço da busca de um consenso; uma estratégia universal, para melhor compreender, comparar e reproduzir diferentes estudos envolvendo expressão relativa. A utilização de dados de transcriptoma global combinada com ferramentas de alinhamento já foi descrita para encontrar genes de referência em gêneros pode ser utilizada como ponto de partida para encontrar genes conservados entre grupos taxonômicos mais abrangentes (LI et al., 2015).

## 8. REFERÊNCIAS

ANDERSEN, C. L.; JENSEN, J. L.; ØRNTOFT, T. F. NormFinder\rNormalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. **Cancer Res**, v. 64, n. 15, p. 5245–5250, 2004.

ANDERSEN, C. L.; JENSEN, J. L.; ØRNTOFT, T. F. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data : A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization , Applied to Bladder and Colon Cancer Data Sets. p. 5245–5250, 2004.

BONASERA, J. M.; ASSELIN, J. A. E.; BEER, S. V. Identification of bacteria pathogenic to or associated with onion (*Allium cepa*) based on sequence differences in a portion of the conserved gyrase B gene. **Journal of Microbiological Methods**, v. 103, p. 138–143, 2014.

BOULTER, N. et al. A simple, accurate and universal method for quantification of PCR. **BMC biotechnology**, v. 16, n. 1, 2016.

BROSCH, R. et al. A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. **Proceedings of the National Academy of Sciences**, v. 99, n. 6, p. 3684–3689, 2002.

BUSTIN, S. A. Absolute quantification of mrna using real-time reverse transcription polymerase chain reaction assays. **Journal of Molecular Endocrinology**, v. 25, n. 2, p. 169–193, 2000.

BUSTIN, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): Trends and problems. **Journal of Molecular Endocrinology**, v. 29, n. 1, p. 23–39, 2002.

BUSTIN, S. A. et al. Quantitative real-time RT-PCR - A perspective. **Journal of Molecular Endocrinology**, v. 34, n. 3, p. 597–601, 2005.

BUSTIN, S. A. et al. The MIQE guidelines:Minimum Information for publication of quantitative real-time PCR experiments. **Clinical Chemistry**, v. 55, n. 4, p. 611–622, 2009.

BUSTIN, S. A. et al. The need for transparency and good practices in the qPCR literature. **Nature Methods**, v. 10, n. 11, p. 1063–1067, 2013.

CARVALHO, D. M. et al. Reference genes for RT-qPCR studies in *Corynebacterium pseudotuberculosis* identified through analysis of RNA-seq data. **Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology**, v. 106, n. 4, p. 605–614, 2014.

CHAOUACHI, M.; BÉRARD, A.; SAÏD, K. Relative quantification in seed GMO analysis: State of art and bottlenecks. **Transgenic Research**, v. 22, n. 3, p. 461–476, 2013.

CHERVONEVA, I. et al. Selection of optimal reference genes for normalization in quantitative RT-PCR. **BMC Bioinformatics**, v. 11, p. 253, 2010.

CHIMARA, E.; FERRAZOLI, L.; LEÃO, S. C. Mycobacterium tuberculosis complex differentiation using gyrB-restriction fragment length polymorphism analysis. **Memorias do Instituto Oswaldo Cruz**, v. 99, n. 7, p. 745–748, 2004.

CHUNG, H. J.; BANG, W.; DRAKE, M. A. Comprehensive Stress Response of *Escherichia coli* in Food Science and Food Safety. v. 5, 2006.

COLE, S. T. et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. **Nature**, v. 393, n. 6685, p. 537–544, 1998.

COSCOLLA, M. et al. Novel *Mycobacterium tuberculosis* complex isolate from a wild chimpanzee. **Emerging Infectious Diseases**, v. 19, n. 6, p. 969–976, 2013.

CUSICK, K. D. et al. Selection and evaluation of reference genes for reverse transcription-quantitative PCR expression studies in a thermophilic bacterium grown under different culture conditions. **PLoS ONE**, v. 10, n. 6, p. 1–23, 2015.

DEAN, B.; UDAWELA, M.; SCARR, E. Validating reference genes using minimally transformed qPCR data: findings in human cortex and outcomes in schizophrenia. **BMC Psychiatry**, v. 16, n. 1, p. 154, 2016.

FREEMAN, W. M.; WALKER, S. J.; VRANA, K. E. Quantitative RT-PCR: Pitfalls and potential. **BioTechniques**, v. 26, n. 1, p. 112–125, 1999.

GACHON, C.; MINGAM, A.; CHARRIER, B. Real-time PCR: What relevance to plant studies? **Journal of Experimental Botany**, v. 55, n. 402, p. 1445–1454, 2004.

GALISA, P. S. et al. Identification and validation of reference genes to study the gene expression in *Gluconacetobacter diazotrophicus* grown in different carbon sources using RT-qPCR. **Journal of Microbiological Methods**, v. 91, n. 1, p. 1–7, 2012.

GUPTA, R. S. The phylogeny of proteobacteria: Relationships to other eubacterial phyla and eukaryotes. **FEMS Microbiology Reviews**, v. 24, n. 4, p. 367–402, 2000.

HEID, C. A. et al. Real time quantitative PCR. **Genome Research**, v. 6, n. 10, p. 986–994, 1996.

HOLT, J. H. et al. Bergey's manual of determinative bacteriology ninth edition. **European journal of paediatric neurology: EJPN: official journal of the European Paediatric Neurology Society**, v. 13, n. 6, p. 560, 1994.

HUARD, R. C. et al. Novel genetic polymorphisms that further delineate the phylogeny of the *Mycobacterium tuberculosis* complex. **Journal of Bacteriology**, v. 188, n. 12, p. 4271–4287, 2006.

HUGGETT, J. et al. Real-time RT-PCR normalisation; strategies and considerations. **Genes and immunity**, v. 6, n. 4, p. 279–84, 2005.

HUGGETT, J. F.; O'GRADY, J.; BUSTIN, S. QPCR, dPCR, NGS - A journey. **Biomolecular Detection and Quantification**, v. 3, n. September 2005, p. A1–A5, 2015.

JOHNSON, G. et al. Quantitative Real-Time PCR. v. 1160, 2014.

KASAI, H.; EZAKI, T.; HARAYAMA, S. Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences. **Journal of Clinical Microbiology**, v. 38, n. 1, p. 301–308, 2000.

KERSTERS, K. et al. Proteobacteria. **Encyclopedia of Life Sciences**, 2006.

KOZERA, B.; RAPACZ, M. Reference genes in real-time PCR. **Journal of Applied Genetics**, v. 54, n. 4, p. 391–406, 2013.

KUBISTA, M. et al. The real-time polymerase chain reaction. **Molecular Aspects of Medicine**, v. 27, n. 2–3, p. 95–125, 2006.

LI, S. et al. Genome-wide identification and characterization of reference genes with different transcript abundances for *Streptomyces coelicolor*. **Scientific reports**, v. 5, n. August, p. 15840, 2015.

LIU, J. et al. Evaluation of *Clostridium ljungdahlii* DSM 13528 reference genes in gene expression studies by qRT-PCR. **Journal of Bioscience and Bioengineering**, v. 116, n. 4, p. 460–464, 2013a.

LIU, J. et al. Evaluation of *Clostridium ljungdahlii* DSM 13528 reference genes in gene expression studies by qRT-PCR. **Journal of Bioscience and Bioengineering**, v. 116, n. 4, p. 460–464, 2013b.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and. **Methods**, v. 25, p. 402–408, 2001.

MOSSBERG, K. et al. Normalization of qPCR in platelets – YWHAE a potential generic reference gene. [Http://Dx.Doi.Org/10.1080/09537104.2016.1180349](http://dx.doi.org/10.1080/09537104.2016.1180349), v. 7104, n. May, 2016.

MULLIS, K. et al. Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction. **Cold Spring Harbor Symposia on Quantitative Biology**, v. 51, n. 1, p. 263–273, 1986.

MULLIS, K.; FERRE, F.; GIBBS, R. The polymerase chain reaction. **NOBEL LECTURES IN CHEMISTRY ...**, v. 28, n. 1–4, p. 44–50, 1994.

NESTOROV, J. et al. Gene Expression Studies: How to Obtain Accurate and Reliable Data by Quantitative Real-Time RT PCR. **Journal of Medical Biochemistry**, v. 32, n. 4, p. 325–338, 2013a.

NESTOROV, J. et al. Gene Expression Studies: How to Obtain Accurate and Reliable Data by Quantitative Real-Time RT PCR / IZUČAVANJE EKSPRESIJE GENA: KAKO DOBITI TAČNE I POUZDANE PODATKE KVANTITATIVNIM RT

PCR-OM U REALNOM VREMENU. **Journal of Medical Biochemistry**, v. 32, n. 4, p. 325–338, 2013b.

PESSOA, D. D. V. et al. Validation of reference genes for RT-qPCR analysis in *Herbaspirillum seropedicae*. **Journal of Microbiological Methods**, v. 127, p. 193–196, 2016.

PFAFFL, M. W. Relative quantification. **Real-time PCR**, p. 63–82, 2004.

POLESE, V. et al. Selection and validation of reference genes for RT-qPCR indicates that juice of sugarcane varieties modulate the expression of C metabolism genes in the endophytic diazotrophic *Herbaspirillum rubrisubalbicans* strain HCC103. **Antonie van Leeuwenhoek**, 2017.

RADONI??, A. et al. Guideline to reference gene selection for quantitative real-time PCR. **Biochemical and Biophysical Research Communications**, v. 313, n. 4, p. 856–862, 2004.

ROBLEDO, D. et al. Analysis of qPCR reference gene stability determination methods and a practical approach for efficiency calculation on a turbot (*Scophthalmus maximus*) gonad dataset. **BMC genomics**, v. 15, n. 1, p. 648, 2014.

ROCHA, D. J. P.; SANTOS, C. S.; PACHECO, L. G. C. Bacterial reference genes for gene expression studies by RT-qPCR: survey and analysis. **Antonie van Leeuwenhoek**, n. i, p. 685–93, 2015.

STACKEBRANDT, E.; MURRAY, R. G. E.; TRUPER, H. G. Proteobacteria classis nov., a Name for the Phylogenetic Taxon That Includes the Purple Bacteria and Their Relatives. **International Journal of Systematic Bacteriology**, v. 38, n. July 1988, p. 321–325, 1988.

TAYLOR, S. et al. A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines. **Methods**, v. 50, n. 4, p. S1–S5, 2010.

TAYLOR, S. C.; MRKUSICH, M. The State of RT-Quantitative PCR: Firsthand Observations of Implementation of Minimum Information for the Publication of Quantitative Real-Time PCR Experiments ( MIQE ). p. 46–52, 2014.

THELLIN, O. et al. Housekeeping genes as internal standards: Use and limits. **Journal of Biotechnology**, v. 75, n. 2–3, p. 291–295, 1999.

THORNTON, B.; BASU, C. Real-time PCR (qPCR) *primer* design using free online software. **Biochemistry and Molecular Biology Education**, v. 39, n. 2, p. 145–154, 2011.

TICHOPAD, A. Standardized determination of real-time PCR efficiency from a single reaction set-up. **Nucleic Acids Research**, v. 31, n. 20, p. 122e–122, 2003.

TOMOYASU, T.; BUKAU, B.; ARSENE, F. The heat shock response of *Escherichia coli*. v. 55, p. 3–9, 2000.

TURRONI, F. et al. Global genome transcription profiling of *Bifidobacterium bifidum* PRL2010 under In Vitro conditions and identification of reference genes for quantitative real-time PCR. **Applied and Environmental Microbiology**, v. 77, n. 24, p. 8578–8587, 2011.

VAN INGEN, J. et al. Characterization of *Mycobacterium orygis* as *M. tuberculosis* complex subspecies. **Emerging Infectious Diseases**, v. 18, n. 4, p. 653–655, 2012.

VANDESOMPELE, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. **Genome biology**, v. 3, n. 7, p. RESEARCH0034, 2002.

VANDRISSE, C. M.; PARKS, A. R.; ESCALANTE-SEMERENA, J. C. A Toxin Involved in *Salmonella* Persistence Regulates Its Activity by Acetylating Its Cognate Antitoxin, a Modification Reversed by CobB Sirtuin Deacetylase. **mBio**, v. 8, n. 3, p. e00708-17, 2017.

VANGUILDER, H. D.; VRANA, K. E.; FREEMAN, W. M. **Twenty-five years of quantitative PCR for gene expression analysis** *BioTechniques*, 2008.

WATERHOUSE, A. M. et al. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189–1191, 2009.

WATTAM, A. R. et al. Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. **Nucleic Acids Research**, v. 45, n. D1, p. D535–D542, 2017.

WEN, S. et al. Validation of reference genes for real-time quantitative PCR (qPCR) Analysis of avibacterium paragallinarum. **PLoS ONE**, v. 11, n. 12, p. 1–14, 2016.

WONG, M. L.; MEDRANO, J. F. Real-time PCR for mRNA quantitation. **BioTechniques**, v. 39, n. 1, p. 75–85, 2005.

XIE, F. et al. Genome-wide functional analysis of the cotton transcriptome by creating an integrated EST database. **PLoS ONE**, v. 6, n. 11, 2011.

XIE, F. et al. miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. **Plant Molecular Biology**, v. 80, n. 1, p. 75–84, 2012.