

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO
EM BIOINFORMÁTICA



Dissertação de Mestrado

**Genômica comparativa de *Francisella
noatunensis subsp. orientalis***

ORIENTADO: **Lucas Amorim Gonçalves**

ORIENTADOR: **Prof. Dr. Vasco Ariston de Carvalho Azevedo**

COORIENTADOR: **Prof. Dr. Henrique César Pereira Figueiredo**

COORIENTADOR: **Prof. Dr. Siomar de Castro Soares**

BELO HORIZONTE

Março - 2015

Lucas Amorim Gonçalves

Genômica comparativa de *Francisella noatunensis subsp. orientalis*

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo programa de Pós-graduação em Bioinformática, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais.

Orientador: Prof. Dr. Vasco Ariston de Carvalho Azevedo

Coorientador: Prof. Dr. Henrique César Pereira Figueiredo

Coorientador: Prof. Dr. Siomar de Castro Soares

BELO HORIZONTE

Março – 2015

DEDICATÓRIA

Dedico esta dissertação à minha esposa Bianca, à minha mãe Terezinha e aos meus irmãos e irmãs Sandra, Márcia, Madson, Meire, Sheila, Robson e Fabiana que sempre estiveram ao meu lado em todos os momentos importantes da minha vida.

Talvez não tenha conseguido fazer o melhor,
mas lutei para que o melhor fosse feito.
Não sou o que deveria ser,
mas graças a Deus,
não sou o que era antes.
(Marthin Luther King)

AGRADECIMENTOS

Agradeço à Universidade Federal de Minas Gerais pela formação acadêmica que me foi oferecida;

Às agências de fomento CNPq, FAPEMIG e CAPES e o Ministério da Pesca e Aquicultura por financiarem este projeto de mestrado;

À coordenação, professores e colegas do curso de Pós-Graduação em Bioinformática da UFMG;

Aos membros da banca por terem aceitado o convite para avaliar este trabalho;

Ao meu orientador prof. Dr. Vasco Azevedo pela oportunidade, por compartilhar sua experiência na elaboração dos trabalhos e por ser uma pessoa sempre disponível;

Ao meu co-orientador prof. Dr. Henrique Figueiredo pela confiança depositada em mim e por ter me recebido junto com a equipe do Aquacen;

Ao meu co-orientador Dr. (e merecido futuro professor da UFTM) Siomar Soares pela confiança e todo apoio dado;

Aos colegas do LGCM por contribuírem para o meu crescimento profissional e por compartilharem diversos momentos agradáveis;

Aos amigos Letícia e Diego por dividirem comigo momentos de alegria e de tristeza durante todo tempo de mestrado;

Aos colegas do Aquacen por terem me recebido de braços abertos;

À Fernanda e Felipe por toda ajuda e informação passada com enorme presteza;

Às meninas da secretaria da bioinformática, especialmente a Sheila, por serem tão eficientes e disponíveis;

À minha família por tudo que ela representa para mim;

E, principalmente, a minha esposa Bianca, que está sempre ao meu lado, sempre carinhosa, paciente e companheira.

A todos vocês, o meu mais sincero “muito obrigado”!

SUMÁRIO

LISTA DE FIGURAS	I
LISTA DE TABELAS	II
LISTA DE ABREVIATURAS	III
RESUMO	1
ABSTRACT	2
COLABORAÇÕES	3
1. CONTEXTUALIZAÇÃO.....	4
2. INTRODUÇÃO	6
2.1. Gênero <i>Francisella</i>	6
2.2. <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	9
2.2.1. Patogênese.....	10
2.3. Tilapicultura.....	12
2.4. Epidemiologia.....	13
3. JUSTIFICATIVA DE REALIZAÇÃO DO TRABALHO.....	18
4. OBJETIVOS.....	19
4.1. Objetivo geral.....	19
4.2. Objetivos específicos.....	19
5. MATERIAL E MÉTODOS.....	20
5.1. Obtenção do DNA genômico para sequenciamento.....	21
5.2. Sequenciamento	22
5.2.1. Sequenciamento utilizando Ion Torrent PGM™.....	23
5.2.2. Sequenciamento utilizando Illumina® MiSeq.....	24
5.3. Montagem.....	25
5.4. Anotação	26
5.4.1. Predição e anotação automática dos genes.....	26
5.4.2. Curadoria Manual.....	26
5.4.3. Curadoria de frameshifts.....	27
5.5. Genômica comparativa.....	29
5.5.1. Plasticidade genômica	30
5.5.2. Representação circular da comparação das sequências genômicas.....	30
5.5.3. Sintenia gênica.....	31
5.5.4. Gegenees.....	31
5.5.5. MLST de 8 genes.....	31
5.5.6. MLST de genoma inteiro ou abordagem gene a gene.....	32

6. RESULTADOS E DISCUSSÃO.	33
6.1. Caracterização dos genomas.	33
6.1.1. Sequenciamento e Montagem.	33
6.1.2. Anotação.	36
6.2. Genômica comparativa.	38
6.2.1. Gegenees.	38
6.2.2. Sintenia gênica.	41
6.2.3. Plasticidade genômica.	44
6.2.4. MLST e MLST de genoma inteiro ou abordagem gene a gene.	50
6.3. Genômica funcional.	54
6.3.1. Estresse oxidativo.	54
6.3.2. Estresse térmico.	58
6.3.3. Estresse ácido.	60
7. CONCLUSÕES.	63
8. PERSPECTIVAS.	64
9. REFERÊNCIAS BIBLIOGRÁFICAS.	65
10. ANEXOS.	72

LISTA DE FIGURAS

Figura 1. Produção de pescado nacional da aquicultura (marinha e continental) durante o período de 1980 a 2010.	4
Figura 2. Baço de tilápia com sinais de infecção granulomatosa severa e crônica ...	9
Figura 3. Tecido de tilápia infectada por <i>Francisella noatunensis</i> subsp. <i>orientalis</i> (FNO).	10
Figura 4. Micrografia eletrônica de macrófagos derivados da porção cranial do rim (HKDM) de tilápia infectados com <i>Francisella noatunensis</i> subsp. <i>orientalis</i> (FNO) LADL 07-285A	11
Figura 5. Gráfico de projetos de genoma por ano e domínio.	16
Figura 6. Desenho experimental das etapas para a caracterização das linhagens de FNO	20
Figura 7. Localização geográfica de fazendas de cultivo de Tilápia do Nilo onde ocorreram surtos de franciselose e data de isolamento das linhagens	22
Figura 8. Mapeamento do conjunto de reads contra o próprio scaffold	25
Figura 9. Fluxo de trabalho da curadoria manual.	27
Figura 10. Fluxo de trabalho da curadoria de frameshifts.	28
Figura 11. Fluxo de trabalho da transferência de anotação.	29
Figura 12. Análise filogenética do gênero <i>Francisella</i>	40
Figura 13. Análise de sintonia das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	43
Figura 14. Análise de plasticidade genômica de <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	45
Figura 15. Relações filogenéticas das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i> e <i>Francisella noatunensis</i> subsp. <i>noatunensis</i>	51
Figura 16. Relações filogenéticas das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	53
Figura 17. Esquema exemplificativo de como as espécies reativas de oxigênio (ROS) produzidas pelas células fagocíticas podem causar danos às bactérias patogênicas.	56
Figura 18. Detecção de peróxido de hidrogênio (H ₂ O ₂) por OxyR.	58
Figura 19. Diferentes condições de estresse encontradas por bactérias patogênicas no ambiente intrafagossômico	61

LISTA DE TABELAS

Tabela 1. Características fisiológicas principais e diferenciais de <i>F. philomiragia</i> subsp. <i>philomiragia</i> , <i>F. noatunensis</i> subsp. <i>noatunensis</i> e <i>F. noatunensis</i> subsp. <i>orientalis</i> . . .	6
Tabela 2. Linhagens do gênero <i>Francisella</i> que possuem genoma completamente sequenciados, montados e anotados disponíveis no NCBI.	8
Tabela 3. Linha do tempo de uma série de avanços tecnológicos e científicos na classificação das doenças infecciosas	14
Tabela 4. Linhagens de FNO que tiveram DNA genômico sequenciado.	21
Tabela 5. Genes housekeeping utilizados para análise de MLST das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	32
Tabela 6. Resumo dos dados de sequenciamento e montagem das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i> isoladas de surtos no Brasil.	35
Tabela 7. Resumo da caracterização dos genomas das linhagens de <i>Francisella noatunensis</i> subsp. <i>orientalis</i> isoladas em diferentes fazendas de cultivo no Brasil . . .	37
Tabela 8. Genes e produtos que compõem as ilhas genômicas preditas para FNO12.	46
Tabela 9. LTRs encontradas na linhagem FNO12.	57
Tabela 10. Genes de choque térmico e frio encontrados em FNO12.	60
Tabela 11. Genes relacionado ao estresse ácido encontrados em FNO12	62

LISTA DE ABREVIATURAS

°C - graus Celsius

ATP - trifosfato de adenosina

BIGSdb - Bacterial Isolate Genome Sequence Database

BLAST - *basic local alignment search tool*

BNDES - Banco Nacional de Desenvolvimento Econômico e Social

BRIG - *blast ring image generator*

CDS - *coding DNA sequence*

CHA - ágar cisteína coração

CR - *Complement receptor*

CSP - proteína de choque frio

DN - *Dichelobacter nodosus*

DNA - ácido desoxirribonucleico

EDTA - *Ethylenediamine tetraacetic acid*

FAO - *Food and Agriculture Organization of the United Nations*

Fe₂⁺ - íon ferroso

Fe-S - ferro-enxofre

FNO - *Francisella noatunensis* subsp. *orientalis*

FNN - *Francisella noatunensis* subsp. *noatunensis*

FPI - *Francisella Pathogenicity Island*

FPP - *Francisella philomiragia* subsp. *philomiragia*

FSP - *Francisella* sp.

FTH - *Francisella. tularensis* subsp. *holartica*

FTM - *Francisella. tularensis* subsp. *mediasiatica*

FTN - *Francisella tularensis* subsp. *novicida*

FTT - *Francisella. tularensis* subsp. *tularensis*

GEI - *Genomic Island*

G+C - conteúdo G+C do genoma

HKDM - *Head-kidney derived macrophages*

H₂O - água

H₂O₂ - peróxido de hidrogênio

HO* - radical hidroxila

HO₂⁺ - íon dioxigênio hidrogênio

HOCl - ácido hipocloroso

HSP - proteína de choque térmico

IDP - Indoxil fosfato

LTTR - *LysR-type transcriptional regulator*

MLST - Multilocus sequence typing

MPA - Ministério da Pesca e Aquicultura

MR - *Mannose receptor*

MI - Ilhas de metabolismo

NADH - nicotinamida adenina dinucleotídeo

NADPH - nicotinamida adenina dinucleotídeo fosfato

NCBI - *national center for biotechnology information*

NGS - *Next Generation Sequence*

O² - oxigênio

OMS - Organização Mundial de Saúde

ONAG - O-nitrofenil-N-acetil-β-D-glicosaminídeo

PAI - *Pathogenicity island*

PCR - Reação em cadeia da polimerase

PFGE - Pulsed field gel electrophoresis

pH - potencial hidrogeniônico

PIB - Produto interno bruto

PUFA - ácidos graxos poli-insaturados

PNPG - P-nitrofenil-β-D-galactopiranosídeo

RAST - *Rapid Annotations using Subsystems Technology*

ROS - *Reactive Oxygen Species*

RI - Ilhas de resistência

rRNA - RNA ribossômico

SI - Ilhas simbióticas

tRNA - RNA transportador

TC - *Thiomicrospira crunogena*

UFC - Unidades Formadoras de Colônia

VNTR - *Variable number tandem repeat*

wgMLST - *Whole genome multi locus sequence typing*

RESUMO

As espécies de *Francisella* são caracterizadas por serem bactérias intracelulares facultativas que causam uma variedade de infecções em animais terrestres e aquáticos. Nesse gênero, *Francisella noatunensis* subsp. *orientalis* (FNO) apresenta uma grande importância por ser um patógeno emergente de peixes, associado a surtos em fazendas de cultura de tilápia-do-nilo. Apesar de a doença causada por FNO apresentar uma alta taxa de mortalidade durante os surtos e ter sido relatada em diversos países, as relações filogenômicas entre os isolados de diferentes países, a história evolucionária deste patógeno e diversos outros aspectos acerca da sua caracterização genômica ainda são pouco conhecidas. Neste contexto, análises de genômica comparativa foram realizadas afim de auxiliar no maior conhecimento dessa espécie de claro interesse epidemiológico. Para tanto, foram sequenciadas, montadas e curadas manualmente 18 linhagens de *F. noatunensis* subsp. *orientalis* isoladas de surtos em fazendas de cultivo de tilápia-do-nilo no Brasil. Em seguida, foram utilizados os softwares Gegenes, Mauve, BRIG, Gipsy e o módulo Genome Comparator do banco de dados BIGSdb para a realização de análises de filogenia, sintenia, plasticidade genômica e polimorfismo. Em todas essas análises foram incluídas outras subespécies de *Francisella* e linhagens de FNO disponíveis no NCBI. Além disso, foram feitas buscas por genes possivelmente envolvidos na resposta a estresses ambientais. Desse modo, foi observado que as linhagens de FNO são altamente clonais e que provavelmente a chegada da doença ao país se deu pela entrada de uma única linhagem. Foram também preditas oito ilhas genômicas, sendo seis consideradas ilhas de patogenicidade. Por fim, a busca por genes potencialmente envolvidos na resposta a estresses ambientais indicou uma série de genes candidatos a um estudo mais aprofundado para o desenvolvimento de drogas.

Palavras chave: *Francisella noatunensis* subsp. *orientalis*, genômica comparativa, tilápia-do-nilo e franciselose.

ABSTRACT

The *Francisella* species are intracellular facultative bacteria that cause a variety of infections in terrestrial and aquatic animals. In this genus, *F. noatunensis* subsp. *orientalis* (FNO) has a remarked importance as an emergent fish pathogen associated with outbreaks in Nile tilapia farms. Although the disease caused by FNO presents a high mortality rate during outbreaks and has been reported in several countries, the phylogenomic relationships among isolates from different countries, the evolutionary history of this pathogen and other aspects still are poorly characterized. In this context, comparative genomic analyzes were performed in order to assist in better understanding this species of great epidemiologic interest. Thus, 18 strains of *F. noatunensis* subsp. *orientalis* isolated from outbreaks in Brazilian Nile tilapia farms were sequenced, assembled and manually curated. Then, the Gegenes, Mauve, BRIG and Gipsy softwares, and the Genome Comparator module of the BIGsdb database were used to perform phylogeny, synteny, genomic plasticity and polymorphisms analyzes. Other *Francisella* subspecies and FNO strains available in the NCBI were included in all of these analyzes. Additionally, searches were made for genes possibly involved in the response to environmental stresses. Therefore, it was observed that the FNO strains are highly clonal and that the arrival of the disease in Brazil is probably due to the entrance of a single strain. Eight genomic islands were also predicted, six being considered pathogenicity islands. Finally, the search for genes potentially involved in the response to environmental stresses indicated a number of candidate genes for a more detailed study for the development of drugs.

Keywords: *Francisella noatunensis* subsp. *orientalis*, comparative genomics, Nile tilapia and francisellosis.

COLABORAÇÕES

O presente estudo foi feito no âmbito de uma colaboração entre o Laboratório de Genética Celular e Molecular (LGCM) sediado no Instituto de Ciências Biológicas (ICB) e o Laboratório Oficial do Ministério da Pesca e Aquicultura (AQUACEN - MPA) sediado na Escola de Veterinária – ambos os institutos da Universidade Federal de Minas Gerais (UFMG). Essas colaborações foram estabelecidas a partir dos interesses destes grupos em promover o desenvolvimento de estudos sobre bactérias de interesse médico veterinário. Este trabalho está sendo desenvolvido sob a orientação do Prof. Dr. Vasco Azevedo (LGCM, UFMG) e a coorientação do Prof. Dr. Henrique César Pereira Figueiredo (AQUACEN, MPA) e Dr. Siomar de Castro Soares. Além disso, este trabalho conta com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG) e do Ministério da Pesca e Aquicultura (MPA).

1. CONTEXTUALIZAÇÃO

Diversas fontes se divergem quanto a origem da aquicultura, mas é consenso que a mesma teve início há mais de dois mil anos com a criação de peixes que eram a base da alimentação, principalmente, de civilizações orientais. O cultivo controlado diminuía a incerteza da pesca; contudo, com o passar do tempo, esta atividade deixou de ser algo para a subsistência e se tornou uma atividade econômica com importante destaque para países asiáticos, principal mercado consumidor. Atualmente, segundo o Ministério da Pesca e Aquicultura (MPA; 2015), a última abrange uma variedade de especialidades, como piscicultura, malacocultura, carcinicultura, entre outros, mostrando uma diversificação e crescimento da atividade.

Segundo a FAO (do inglês Food and Agriculture Organization of the United Nations), alimentar a população mundial – que ultrapassou a quantidade de 7 bilhões em 2011 e atingirá 8,3 bilhões no ano de 2030 – é um desafio. Desse modo, a demanda por proteína animal aumentou e continuará aumentando, por algum tempo, de forma vertiginosa, sendo que o Brasil tem um papel de destaque nesse contexto uma vez que é um dos principais produtores de carne e tem enorme potencial de crescimento na indústria da pesca.

Apesar de ainda pouco estruturada e necessitando de avanços e investimentos na indústria da pesca, o país dispõe de um extenso litoral e a maior reserva de água doce do mundo, e vê sua participação no mercado e produção de pescados crescer (Figura 1).

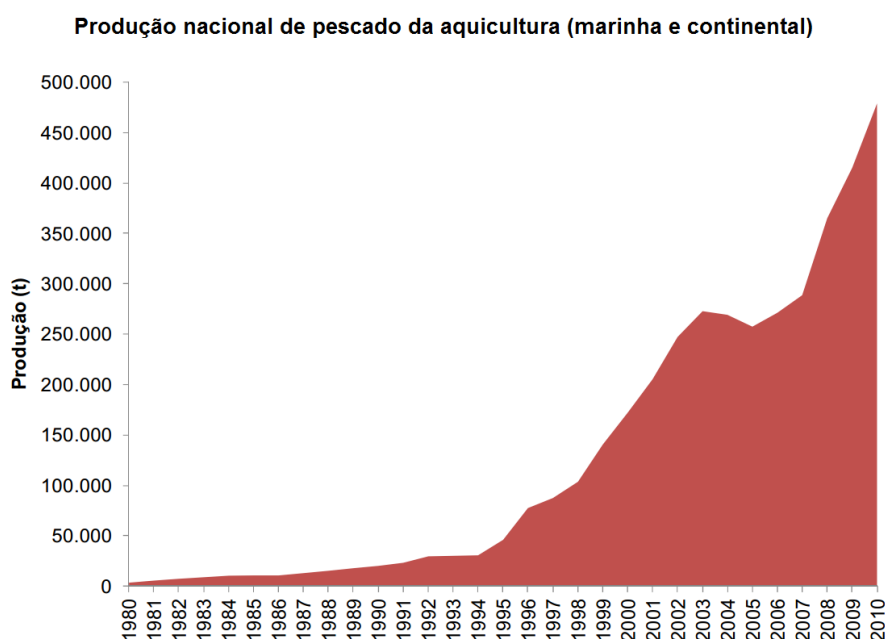


Figura 1: Produção de pescado nacional da aquicultura (marinha e continental) durante o período de 1980 a 2010. **ADAPTADO DE:** Boletim Estatístico da Pesca e Aquicultura - MPA, 2012.

De acordo com dados preliminares apurados pelo MPA para o ano de 2013, o país produz cerca de 2 milhões de toneladas de pescado que é a soma da produção da aquicultura, representada no gráfico acima, e da pesca extrativa. O PIB (Produto interno bruto) da atividade é em torno de 5 bilhões e envolve 800 mil profissionais da pesca e aquicultura, proporcionando 3,5 milhões de empregos diretos e indiretos.

Segundo Callou (2010), essas comunidades pesqueiras em sua maioria habitam o litoral e compõem uma população humilde quanto a formação técnica, posses e tem na atividade a única forma de obtenção de renda. No intuito de reverter esse quadro, o governo brasileiro tem, nos últimos anos, por meio do MPA e BNDES (Banco Nacional de Desenvolvimento Econômico e Social), incentivado uma melhoria quanto a capacitação, infraestrutura e qualidade do produto.

Baseada principalmente nesses incentivos, a produção de tilápias e camarões ajudou a aquicultura brasileira a ser a segunda maior da América do Sul no fim da última década, com uma produção de quase 133 mil de toneladas de tilápias. Somente na primeira década dos anos 2000, a produção de tilápia cresceu 105% e ainda possui condições favoráveis para um maior incremento desses números.

Entretanto, o aquicultor enfrenta outros desafios inerentes a qualquer tipo de cultivo: a exposição a doenças que trazem perdas ou até a interrupção da produção. Segundo Leal *et al.* (2014), foram relatados surtos de infecções granulomatosas com altas taxas de mortalidade em fazendas de cultivo de tilápia no Brasil. O agente infeccioso em questão é a *Francisella noatunensis subsp. orientalis*, o qual é um patógeno emergente (Soto *et al.* 2009) que já teve surtos reportados em Taiwan, Costa Rica, Indonésia, Reino Unido e Estados Unidos (Leal *et al.* 2014).

Logo, a doença e o patógeno supracitados tem uma distribuição mundial e alto potencial para perdas econômicas na indústria da pesca, despertando assim o interesse para a investigação desse quadro.

2. INTRODUÇÃO

2.1. Gênero *Francisella*

Em 1922, o bacteriologista americano Edward Francis (1872-1957) descreveu pela primeira vez a bactéria causadora da tularemia em humanos, *Francisella tularensis*, a qual é a espécie mais estudada do gênero *Francisella* (Sridhar *et al.*, 2012; Soto *et al.*, 2013).

Este gênero (Filo: Proteobacteria; Classe: Gammaproteobacteria; Ordem: Thiotrichales; Família: Francisellaceae) é caracterizado por bactérias Gram-negativas pleomórficas de 0,5-1,5 µm, aeróbias estritas e imóveis. Elas também são negativas para a redução de nitrato; fermentação de adonitol, arabinose, celobiose, esculina, galacturonato, glicuronato, rafinose, malonato, manitol, melibiose, rafinose, ramnose, palatinose e 5-cetogliconato; presença de atividade de C14 lipase, cisteína arilamidase, para-fenilalanina desaminase, tetrionato redutase, tripsina, urease, valina arilamidase, α-fucosidase, α-galactosidase, α-manosidase, α-quimotripsina e β-glicuronidase; e produção de ácido a partir de lactose. Além disso, essas bactérias são positivas para a presença de atividade de C4 esterase, C8 esterase, fosfatase ácida, fosfatase alcalina, lipase, naftol-AS-BI-fosfohidrolase e β-lactamase; e produção de ácido a partir de maltose (Birkbeck *et al.*, 2011) (Tabela 1).

Tabela 1: Características fisiológicas principais e diferenciais de *F. philomiragia* subsp. *philomiragia*^{*}, *F. noatunensis* subsp. *noatunensis*^{*} e *F. noatunensis* subsp. *orientalis*^{*}. ADAPTADO DE: Birkbeck *et al.*, 2011.

Característica	<i>F. philomiragia</i>	<i>F. noatunensis</i>	<i>FNO</i>
Requerimento de cisteína p/ o crescimento	-	+	+
Atividade:			
Citocromo oxidase	+	-	-
Catalase	F	F	F
Indol	F	F	F
Produção de ácido a partir de:			
Sacarose	+	-	-
D-glicose	+	+	-
β-galactosidade	+	-	-
Deteção de atividade enzimática de:			
O-nitrofenil-N-acetil-β-D-glicosaminídeo (ONAG)	+	-	-
P-nitrofenil-β-D-galactopiranosídeo (PNPG)	+	-	-
Leucina arilamidase	+	-	-
N-acetil-β-glicosaminidase	+	-	-
Indoxil fosfato (IDP)	+	-	+

O sinal +: é espécie reativa para a característica avaliada. O sinal -: não é espécie reativa para a característica avaliada. A sigla F significa baixa reatividade segundo Mikalsen & Colquhoun (2010).

Microrganismos intracelulares facultativos, as bactérias desse gênero são capazes de infectar uma variedade de animais terrestres e aquáticos, podendo ser categorizadas em metabolicamente competentes, que são capazes de se desenvolver independentemente de um hospedeiro (microrganismos de vida livre); e metabolicamente incompetentes, que dependem de um hospedeiro para se desenvolverem. Além disso, podem ser classificadas em não virulentas, moderadamente virulentas e virulentas (Sridhar *et al.*, 2012; Soto *et al.*, 2013).

Várias espécies de *Francisella* têm sido detectadas em amostras ambientais, como simbioses em organismos, e em humanos (Niebylski *et al.* 1997; Barns *et al.* 2005; Kugeler *et al.* 2008) e peixes infectados. Até recentemente, o gênero *Francisella* consistia em apenas duas espécies: *F. tularensis subsp. tularensis* e *F. philomiragia subsp. philomiragia*. Porém, novas espécies e novas linhagens foram isoladas, como *F. noatunensis subsp. noatunensis* (patógeno de bacalhau) e *F. noatunensis subsp. orientalis* (syn. *F. asiatica*) (Ottem *et al.*, 2009), sendo esta reconhecida como um dos patógenos mais importantes de tilápia (*Oreochromis* spp.) cultivada (Soto & Revan, 2012).

Atualmente, se considerarmos apenas linhagens com genomas completamente sequenciados, montados e anotados esse número chega a 19 exemplares (Tabela 2) e esse valor aumenta para 237 se buscarmos a quantidade de "bioprojects" (coleções de dados biológicos) depositados no NCBI (do inglês, "National Center for Biotechnology Information" - <http://www.ncbi.nlm.nih.gov>. Último acesso: 20/02/2015).

Tabela 2: Linhagens do gênero *Francisella* que possuem genoma completamente sequenciados, montados e anotados disponíveis no NCBI

Linhagem	Tam. (Mb)	GC%	rRNA	tRNA	Gene	Pseudogene
FTT NE061598	1892681	32%	10	38	1966	116
FTT SCHU S4	1892789	32%	10	38	1962	119
FTT TI0902	1892744	32%	10	38	1967	114
FTT TIGB03	1968651	32%	10	39	2066	116
FTT WY96-3418	1898476	32%	10	38	1973	119
FTT FSC198	1892616	32%	10	38	1962	116
FTM FSC147	1893886	32%	10	38	1977	254
FTH FSC200	1894157	32%	10	38	1999	164
FTH FTNF002-00	1890909	32%	10	38	1990	177
FTH LVS	1892177	32%	10	38	1993	169
FTH OSU18	1895727	32%	10	38	1997	194
FTH F92	1886888	32%	10	38	1988	221
FSP TX077308	2035931	32%	10	39	1972	39
FPP ATCC 25017	2045775	32%	10	39	1983	34
FTN U112	1910592	32%	10	38	1837	21
FTN 3523	1945310	32%	6	38	1854	-
FTN Fx1	1913619	32%	10	38	1825	32
FNO LADL-07-285A	1858987	32%	10	39	1991	317
FNO Toba04	1847202	32%	4	35	1860	252

FTT: *Francisella. tularensis* subsp. *tularensis*; FTM: *Francisella. tularensis* subsp. *mediasiatica*; FTH: *Francisella. tularensis* subsp. *holartica*; FPP: *Francisella philomiragia* subsp. *philomiragia*; FSP: *Francisella* sp.; FTN.: *Francisella tularensis* subsp. *novicida* e FNO: *Francisella noatunensis* subsp. *orientalis*.

Tal gênero é formado por espécies que têm genomas compostos, em média, por 2 milhões de pares de bases, conteúdo GC de aproximadamente 32%, 46 RNAs estruturais e até 2 ilhas de patogenicidade. Como dito anteriormente, as espécies de *Francisella* podem ser metabolicamente competentes e incompetentes, sendo essa competência relacionada ao número de genes interrompidos (parcialmente conservados ou "frameshifts"). *F. tularensis* subsp. *tularensis*, *F. tularensis* subsp. *holartica*, *F. tularensis* subsp. *mediasiatica* – as quais são patógenos de mamíferos – e *F. noatunensis* subsp. *orientalis* – a qual é patógeno de peixe – são metabolicamente incompetentes e têm um alto número de frameshifts. Por outro lado, *F. philomiragia* subsp. *philomiragia* e *F. tularensis* subsp. *novicida* – espécies de vida livre – são metabolicamente competentes e têm poucos genes interrompidos (Tabela 2) (Sridhar *et al.*, 2012).

Apesar de muitos aspectos microbiológicos e caracterizações bioquímicas e genômicas estarem disponíveis para o gênero *Francisella* os estudos se concentram nas subespécies de *Francisella tularensis* pelo interesse ligado a saúde humana e biossegurança (Oyston *et al.*, 2004; Brudal *et al.*, 2014). Neste contexto, outras espécies como *F. noatunensis* subsp. *orientalis*, patógenos com importância veterinária e econômica (Soto & Revan, 2012), são pouco estudadas.

2.2. *Francisella noatunensis* subsp. *orientalis*

F. noatunensis subsp. *orientalis* (FNO) (syn. *F. asiatica*) é o agente etiológico das infecções piogranulomatosa e granulomatosa em peixes. Nos últimos anos, FNO tem sido a responsável por uma grande quantidade de mortes de tilápia e outras espécies de água-doce cultivadas nos Estados Unidos, Reino Unido, Japão, Taiwan, Jamaica, Costa Rica, Brasil e outras regiões da América Latina (Soto *et al.*, 2012a; 2012b; Leal *et al.* 2014). Todavia, além de infectar espécies cultiváveis importantes como tilápia-do-nylo (*Oreochromis niloticus*), "threeline grunt" (*Parapristipoma trilineatum*) e robalo híbrido (*Morone chrysops* X *Morone saxatilis*), essa bactéria também é capaz de infectar peixes selvagens como guapote tigre (*Parachromis managuensis*) (Soto *et al.*, 2012b).

A doença causada pela FNO, também chamada de franciselose (Leal *et al.*, 2014) pode se apresentar de forma aguda, com poucos sinais clínicos e alta mortalidade, e de forma subaguda à crônica, com sinais clínicos inespecíficos que incluem exoftalmia, anorexia e anemia. Além disso, os órgãos internos do hospedeiro apresentam-se aumentados e com diversos nódulos brancos pouco protuberantes (Figura 2). Ademais, o exame de tecidos afetados revela várias lesões inflamatórias granulomatosas contendo numerosos cocobacilos pequenos (Soto *et al.*, 2012b) (Figura 3). Um fato interessante, porém, é que foi demonstrado que apenas 23 UFCs (Unidades Formadoras de Colônia) desse microrganismo são capazes de causar a morte de tilápias e que ainda menos UFCs são necessárias para causar lesões patológicas significantes no rim e no baço (Soto *et al.*, 2009; 2012b; Soto & Revan, 2012). No entanto, os mecanismos responsáveis pela notável infectividade da FNO em diversos hospedeiros ainda são pouco conhecidos (Soto *et al.*, 2010).

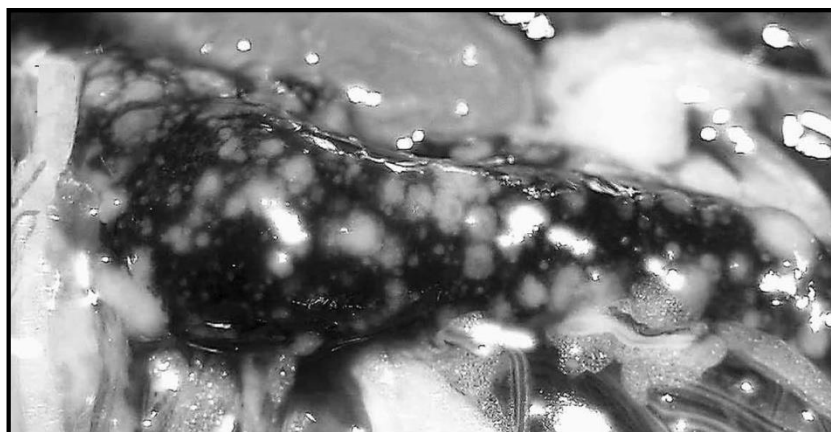


Figura 2: Baço de tilápia com sinais de infecção granulomatosa severa e crônica. **ADAPTADO DE:** Colquhoun & Duodu, 2011.

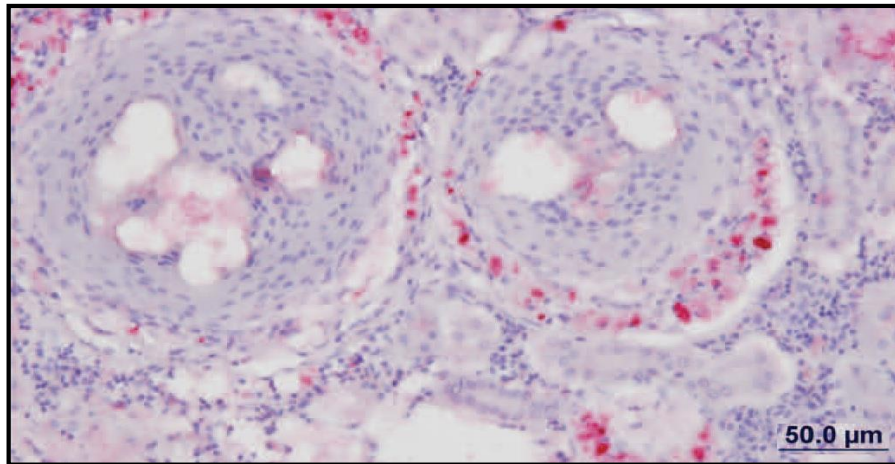


Figura 3: Tecido de tilápia infectada por *Francisella noatunensis subsp. orientalis* (FNO). A cor vermelha mostra a reação positiva para o soro polivalente de camundongo anti-FNO, evidenciando a coloração intensa dentro dos granulomas e nas células adjacentes. **ADAPTADO DE:** Soto *et al.*, 2012a.

2.2.1. Patogênese

Na última década, estudos acerca da virulência de *Francisella* têm aumentado significativamente. Sabe-se que ela é dependente da habilidade do microrganismo de entrar, persistir e se replicar nas células hospedeiras uma vez que as linhagens que perderam essa habilidade são avirulentas (Jones *et al.*, 2014).

De um modo geral, dependendo do seu estado de opsonização, FNO é capaz de interagir com diferentes receptores ao entrar em contato com uma célula fagocítica. Soto *et al.* (2010) demonstraram que a internalização efetiva de FNO por macrófagos derivados da porção cranial do rim (HKDM - do inglês, "head-kidney derived macrophages") foi parcialmente mediada por um componente do soro sanguíneo de tilápia que é sensível ao calor, sendo ele presumivelmente o complemento. Além disso, foi demonstrado que o receptor de manose (MR - do inglês, "mannose receptor") também contribui para a internalização desse microrganismo por HKDM. No entanto, existem evidências de que o receptor do complemento (CR - do inglês, "complement receptor") e o MR não são os únicos receptores envolvidos na internalização de FNO (Soto *et al.*, 2010).

Apesar de não ser bem conhecido o porquê da FNO ser capaz de utilizar diversos receptores para serem internalizadas por células fagocíticas, sabe-se que isso a torna apta a ser internalizada por mais de um tipo celular. A sua internalização por meio do uso de receptores diferentes ativa cascatas de sinalização distintas, induzindo respostas imunológicas diferentes nas células hospedeiras. Isso, por sua vez, indica que FNO desenvolveu múltiplos mecanismos de internalização para usufruir de diferentes

condições ambientais, que parecem ser importantes para as interações subsequentes desse patógeno com o seu hospedeiro (Jones *et al.*, 2014).

Após a sua internalização, FNO parece se alojar num pequeno fagossomo cuja membrana é, posteriormente, rompida, permitindo o escape do microrganismo para o citoplasma da célula hospedeira. Porém, 12 horas após a infecção, a maioria das células bacterianas ainda se encontram em fagossomos, mostrando que são necessários estudos que elucidem a localização celular de FNO em tempos posteriores de infecção (Figura 4) (Soto *et al.*, 2010). O mecanismo molecular que permite que FNO escape do fagossomo ainda não é conhecido, porém diversos estudos demonstraram que a ruptura de genes pertencentes à ilha de patogenicidade de *Francisella* (FPI - do inglês, "*Francisella Pathogenicity Island*") torna bactérias desse gênero incapazes de executar tal ação (Jones *et al.*, 2014).

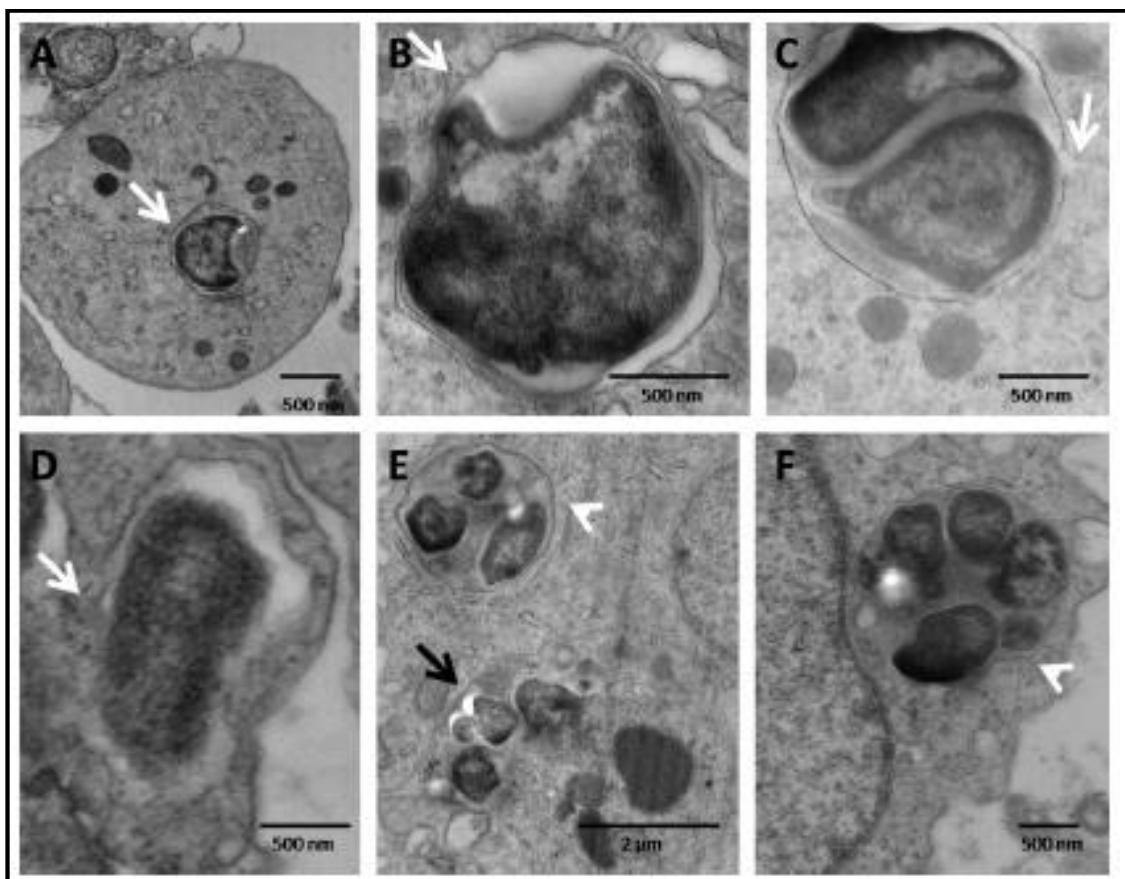


Figura 4: Micrografia eletrônica de macrófagos derivados da porção cranial do rim (HKDM) de tilápia infectados com *Francisella noatunensis subsp. orientalis* (FNO) LADL 07-285A. (A e B): Logo após a infecção, a bactéria se encontra em fagossomos (setas brancas) dentro dos macrófagos. (C e D): Seis a 12 horas após a infecção, o rompimento da membrana do fagossomo (setas brancas) permite que a bactéria acesse o citoplasma das células hospedeiras. (E e F): Doze horas após a infecção, várias células bacterianas são encontradas em vacúolos espaçosos (cabeça de seta), sendo que algumas delas parecem ter escapado para o citoplasma do macrófago (seta preta). **ADAPTADO DE:** Soto *et al.*, 2010.

Alguns genes que codificam fatores bacterianos essenciais para o crescimento citosólico foram identificados em *F. tularensis* e *F. novicida*, sendo eles envolvidos na biossíntese de purinas (*purMCD*) ou de uracila (*pyrF*); no transporte de asparagina (gene ortólogo de *ansP*); e na replicação bacteriana, como *dipA* e *ripA* (Jones *et al.*, 2014).

No entanto, embora a morte celular possa funcionar como um mecanismo de defesa, privando o microrganismo adaptado ao ambiente intracelular do seu habitat preferencial, foi demonstrado por Soto *et al.* (2010) que a infecção por FNO é pró-apoptótica uma vez que, 36 horas após a infecção, HKDM apresentaram um aumento significativo na atividade das enzimas caspase 3 e caspase 7, as quais são marcadores da apoptose. Assim, acredita-se que o evento de apoptose induzido por FNO permite que a bactéria se replique dentro das células hospedeiras, permitindo que elas escapem para o ambiente extracelular quando os recursos celulares são consumidos (Soto *et al.*, 2010).

2.3. Tilapicultura

Estudos detalhados sobre a caracterização genômica e patogenia de parasitas de peixe são de extrema importância, uma vez que eles apresentam o principal entrave para a expansão da tilapicultura no país (Mian *et al.*, 2009; Pereira *et al.* 2010). A carne de pescado é um importante alimento da dieta das populações de muitos países, é fonte importante de emprego, lucro e renda (Gonçalves *et al.*, 2008; Santos, 2011).

Devido às muitas qualidades nutricionais, como presença de vitaminas A e D, íons cálcio e fosforo, alto teor proteico e ácido graxos insaturados como ômega 3 e 6 presentes em algumas espécies associadas à redução da incidência de doenças cardiovasculares, houve um crescimento no consumo de pescados (Gonçalves *et al.*, 2009; Santos, 2011; Silva *et al.*, 2013).

Proveniente da África, a tilápia do Nilo é a espécie de peixe que possui a maior relevância e produção na aquicultura nacional (IBGE, 2013). Isso se deve em grande parte a rusticidade, resistência a doenças e variações de temperatura e oxigênio no ambiente apresentada pela espécie. Adicionalmente, ela se reproduz facilmente, apresenta rápido crescimento e baixo custo de produção, além de elevado valor proteico e qualidade nutricional (Rodrigues, 2007).

A popularização da tilapicultura no Brasil levou a um aumento no número de produtores e intensificação dos sistemas de produção. O sistema mais tradicional e antigo de produção praticado no país é a criação em tanques escavados, que apresenta disponibilidade de alimento natural associado ao maior espaço físico, conferindo a esta

espécie maior conforto e resistência a enfermidades. Já o sistema intensivo que utiliza tanques-rede é caracterizado pela alta renovação de água e por criações de altas densidades (Bozano e Cyrino, 1999). Porém, nessa estratégia as tilápias são submetidas a estresses derivados do manejo e estresse físico.

Desse modo, o manejo adequado e a qualidade da água influencia de maneira marcante a resistência do animal à doenças e patógenos oportunistas que levam à queda na produção e perdas econômicas (Kubitza, 2005). Isso exige um conhecimento tanto dos processos de produção como a elaboração de estratégias para o controle e prevenção de surtos de doenças causadas principalmente por parasitas, o que evidencia a importância de estudos epidemiológicos.

2.4. Epidemiologia

Segundo a Organização Mundial de Saúde (OMS), epidemiologia é o estudo da distribuição e dos determinantes de estados ou eventos relacionados à saúde (incluindo a doença), e a aplicação de tal estudo no controle de doenças ou outros problemas de saúde. Vários métodos podem ser usados para conduzir investigações epidemiológicas: estudos de vigilância e estudos descritivos podem ser usados para estudar a distribuição, enquanto estudos analíticos são usados na descoberta de fatores determinantes da doença (disponível no endereço eletrônico <http://www.who.int/topics/epidemiology/en/>. Acessado em: 08/03/2015).

Historicamente, a vigilância epidemiológica de doenças infecciosas procurou transformar os dados sobre as populações infectadas em inferências sobre onde e como ocorre a propagação de doenças infecciosas. O exemplo mais clássico é o surto de cólera em Londres no século 19 em que John Snow mapeou os locais de casos da doença e inferiu que o surto foi devido à água contaminada de uma bomba localizada na Broad Street, mesmo antes da identificação do agente etiológico *Vibrio cholerae*. A coleta de algumas informações críticas como a localização geográfica, idade, sexo, características clínicas e data de diagnóstico são alguns dos metadados utilizados para criar e testar hipóteses sobre a transmissão e susceptibilidade da população à doença (Grad & Lipsitch, 2014).

Tendo como referência, principalmente, as contribuições da bioinformática para estudos epidemiológicos, pode-se dividir a epidemiologia em duas fases: pré-genômica e pós-genômica (Croucher et al. 2013) (Tabela 3).

Tabela 3: Linha do tempo de uma série de avanços tecnológicos e científicos na classificação das doenças infecciosas. **ADAPTADO DE:** Grad & Lipsitch, 2014.

Data	Fato / marco	Aplicações
1670	Invenção do Microscópio por Leeuwenhoek	Visualização de bactérias e protozoários
1850	Identificação da febre puerperal como infecciosa e intervenções implementadas por Semmelweis	Controle de infecção hospitalar motivada pela crescente compreensão da etiologia microbiana
1864	Transmissão de cólera por água, comprovado por John Snow	Medidas de prevenção da síndrome infecciosa
1890	Comprovação da origem parasitária da malária e transmissão pelo mosquito	Controle vetorial
1890	Identificação de etiologias microbianas para tuberculose e anthrax; Postulados de Koch	Diagnósticos e terapias direcionadas para o patógeno identificado
1910 - 1950	Taxonomia fenotípica de subespécies: sorotipagem e tipagem de fago	Associação de subespécies específicas com prognóstico, e resistência a drogas
1944	Descoberta do DNA como material genético	Base para ferramentas de genotipagem para epidemiologia molecular
1970	Enzimas de restrição	Base para metodologias de polimorfismos de comprimento de fragmentos de restrição, por exemplo (PFGE - eletroforese em gel de campo pulsado)
1975 - 1985	Sequenciamento de DNA pelo método de Sanger; Reação em cadeia da polimerase (PCR)	Base para metodologias de números variáveis de repetições em tandem (VNTR) e tipagem de sequência multilocus (MLST)
2000 - Hoje	NGS (Next Generation Sequence)	Sequenciamento de genomas inteiros

O sombreamento indica a fase pós – genômica e os fatos / marcos e aplicações destacadas em negrito indicam contribuições da bioinformática.

A fase pré-genômica na epidemiologia ficou marcada principalmente por métodos de tipagem microbiana que permitem a caracterização da bactéria e fornecem aos pesquisadores importantes informações para a vigilância de doenças infecciosas, investigação e controle de surtos. Estes métodos oferecem novos entendimentos sobre a patogênese e a história natural de uma infecção e também dentro da genética de populações bacteriana, área de pesquisa que tem um impacto importante nos problemas de saúde tais como o desenvolvimento de vacinas ou novas drogas antimicrobianas, com significantes implicações econômicas e sociais. (Carriço *et al.*, 2013; Croucher *et al.*, 2013; Grad & Lipsitch, 2014).

Os principais métodos moleculares são eletroforese em gel de campo pulsado (PFGE – do inglês *Pulsed-field gel electrophoresis*) e tipagem de sequência multilocus (MLST – do inglês *Multilocus sequence typing*). O primeiro consiste em uma técnica que permite a separação de um pequeno número de fragmentos de restrição de DNA grandes produzidos usando enzimas de restrição para gerar uma impressão digital genética altamente discriminatória (Griffiths *et al.*, 2006). Já o segundo consiste no sequenciamento de fragmentos de 400-500 pares de bases de 7 ou mais genes conservados, ditos housekeeping, que permite a detecção de pequenas variações dentro de uma espécie (Larsen *et al.*, 2012).

Métodos como o PFGE forneceram reprodutibilidade intra- e inter-laboratório necessária para criar os bancos de dados de isolados que podem ser usados para diversos estudos. Isto permitiu que a tipagem bacteriana se estendesse além da investigação de surtos e os resultados eram alocados em bancos de dados locais, usando softwares especializados tais como *BioImage Whole Band Analyzer* e *GelCompar* (Bionumerics). Estes softwares, que integravam gerenciadores de bancos de dados rudimentares e análises de imagem de géis, foram de fato as primeiras ferramentas de bioinformática amplamente adotadas a serem usadas no campo. Isso somado à possibilidade de compartilhar informação usando a internet levou à criação do banco de dados PulseNet (Carriço *et al.*, 2013).

Contudo, o banco de dados para (MLST) obteve um sucesso maior devido à portabilidade inerente dos dados de sequência, independentes de variabilidades inerentes à realização de corridas eletroforéticas. Ao contrário do PulseNet, sites de MLST hospedam bancos de dados acessíveis publicamente onde qualquer laboratório pode submeter dados. A habilidade de compartilhar dados de sequência com facilidade através da internet é uma das características que tornaram o MLST como o método de escolha para identificação clonal e rastreamento para muitas espécies bacterianas. Outra característica importante que contribuiu para o sucesso do MLST foi a possibilidade de inferir padrões de descendência filogenética através da comparação

dos perfis alélicos. Apesar de o MLST ter se tornado o padrão ouro para a vigilância epidemiológica a longo-termo de várias espécies, PFGE continua importante para a detecção de surtos por que este frequentemente apresenta um poder discriminatório maior (Carricho *et al.*, 2013; Maiden *et al.*, 2013).

Com o advento das novas tecnologias de sequenciamento, conhecidas como sequenciadores NGS, iniciou-se a fase pós-genômica (Maiden *et al.*, 2013). Por meio de plataformas de sequenciamento como 454 GS FLX system (Roche), Illumina GA IIx (Illumina), HeliScope (Helicos), SOLiD 5500 XL system (ABI), PacBio RS system (Pacific Biosciences) e a Personal Genome Machine - Ion Torrent (Life Technologies) o volume de dados genômicos e informações aumentou drasticamente (Kato, 2009) (Figura 5). O sequenciamento genômico não era um processo recente, pois na década de 70 Frederick Sanger desenvolveu a tecnologia de sequenciamento de DNA baseada no método de terminação de cadeia e Allan Maxam e Walter Gilbert elaboraram a tecnologia de sequenciamento químico de degradação de bases (Kaur & Malik, 2013). Porém, somente com a chegada dos sequenciadores de nova geração é que se obteve a possibilidade de estudar sequências de genomas completos ou quase completos de milhares de linhagens individuais de forma rápida e barata (Metzker, 2005; Harismendy *et al.* 2009; El-Metwally *et al.*, 2013).

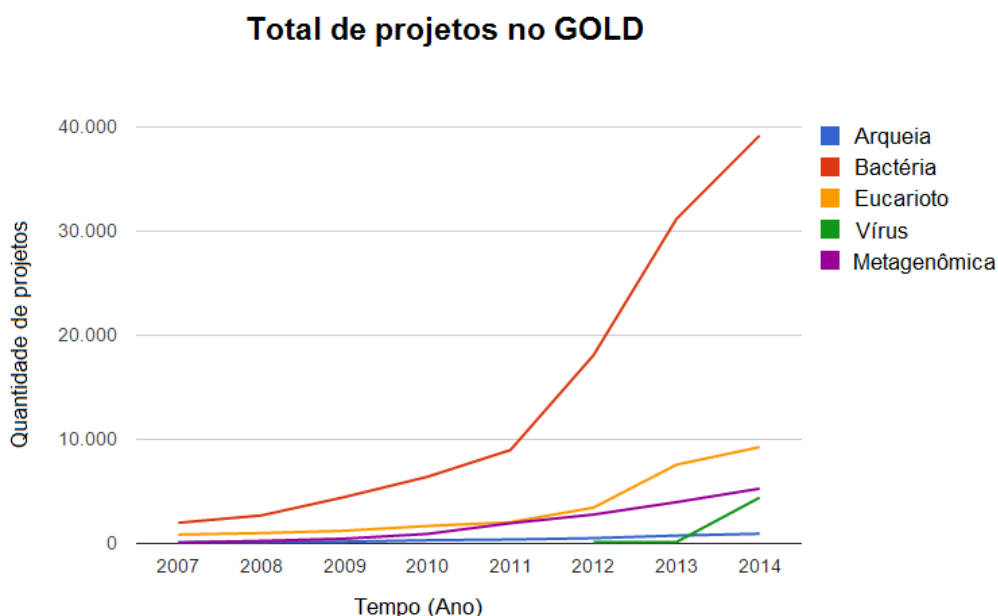


Figura 5: Gráfico de projetos de genoma por ano e domínio. (Adaptado de <http://genomesonline.org/>. Último acesso: 23/01/2015).

O sequenciamento completo de genomas é uma progressão histórica no estudo das doenças para a epidemiologia, pois a melhora na caracterização de patógenos e

compreensão mais profunda da biologia dos mesmos impulsionou o desenvolvimento de técnicas de diagnóstico, vacinas e terapias (Grad & Lipsitch, 2014). Contrastando com a maioria dos métodos de tipagem, o sequenciamento de genoma inteiro facilita a identificação direta das perdas e ganhos de gênicos que podem desempenhar um papel na evolução de uma espécie (Parkhill & Wren, 2011).

No entanto, apesar de tantos avanços e grandes promessas dessas metodologias, ainda não está claro como dados de genomas completos de patógenos bacterianos serão compartilhados e usados para a vigilância de populações bacterianas e as possíveis aplicações em saúde pública. Nesse contexto, o BIGSdb (do inglês Bacterial Isolate Genome Sequence Database) se apresenta como uma opção para manusear dados NGS de genomas microbianos e executar análises focadas em métodos de tipagem MLST estendidos, que podem compreender milhares de genes, e também em outras metodologias de análises de populações. Na estratégia de whole genome MLST (wgMLST) realizada pelo software BIGSdb, é possível não apenas discriminar linhagens isoladas de regiões diferentes, como também diferenciar populações diferentes de uma dada espécie em um mesmo paciente. Este poder discriminatório do BIGSdb demonstra o potencial que as análises bioinformáticas de genoma inteiro apresentam para estudos futuros (Jolley & Maiden, 2010; Maiden *et al.*, 2013).

Desse modo, a integração dos dados já estabelecidos dos métodos de tipagem microbiana, dos bancos de dados genômicos e epidemiológicos e de dados de NGS serão a próxima fronteira na epidemiologia bacteriana. Uma vez que NGS se tornou amplamente adotado, o desenvolvimento de software que analise a informação de diferentes fontes de dados será uma chave para a síntese de conhecimentos disponíveis (Carriço *et al.*, 2013).

3. JUSTIFICATIVA DE REALIZAÇÃO DO TRABALHO

Além de ser um patógeno com distribuição mundial, a franciselose é uma das principais causas de perdas econômicas para a indústria da pesca e aquicultura (Soto *et al.* 2009; 2012c), inclusive com surtos em fazendas de cultivo de tilápia no Brasil (Leal *et al.* 2014). Apesar da evidente importância, trata-se de um organismo não muito estudado, necessitando, assim, de estudos mais aprofundados para a compreensão da sua biologia (Soto *et al.* 2012c).

Estudos de genômica comparativa podem identificar fatores importantes para a patogenicidade e sobrevivência de FNO tanto no meio ambiente quanto em seu hospedeiro. Análises de bioinformática podem elucidar a fonte de contaminação dos cultivos de tilápia no Brasil e auxiliar na determinação da história evolutiva e diversidade genética de diferentes isolados de FNO. Isso, por sua vez, constitui uma abordagem que irá, possivelmente, contribuir para que, num futuro próximo, estratégias de prevenção, controle e tratamento possam ser uma realidade e que perdas no cultivo, principalmente, de tilápia sejam mitigados ou até evitados.

4. OBJETIVOS

4.1. Objetivo geral

Caracterizar os genomas de 18 linhagens *Francisella noatunensis* subsp. *orientalis* isoladas de diferentes fazendas de cultivo de tilápia do Nilo e realizar análises comparativas entre elas, incluindo genomas sequenciados da mesma espécie disponíveis no NCBI.

4.2. Objetivos específicos

- Sequenciar, montar, anotar e curar manualmente os genomas das 18 linhagens de *Francisella noatunensis* subsp. *orientalis*;
- Caracterizar e anotar os genomas quanto: ao conteúdo G+C, número de regiões codificantes de proteína, à quantidade de RNAs ribossomais e transportadores;
- Analisar a similaridade e sintonia entre os diferentes genomas, incluindo os completos já depositados no banco de dados do NCBI;
- Evolução das linhagens de FNO em relação ao seu gênero;
- Analisar a distribuição geográfica das linhagens de FNO do Brasil e sua correlação com a epidemiologia da doença;
- Identificar possíveis ilhas genômicas e correlacionar a plasticidade do seu genoma com outras espécies do gênero.
- Identificar possíveis genes relacionados a mecanismos de resistência a estresses ambientais.

5. MATERIAL E MÉTODOS

Para um melhor entendimento do trabalho foi feito um desenho experimental (Figura 6) que consisti em três etapas: parte I, na qual foi feita a caracterização dos genomas sequenciados isolados de surtos ocorridos no Brasil; parte II, em foram realizadas análises comparativas entre as linhagens de interesse e parte III na qual foram feitas buscas por genes possivelmente envolvidos na resposta a estresses ambientais.

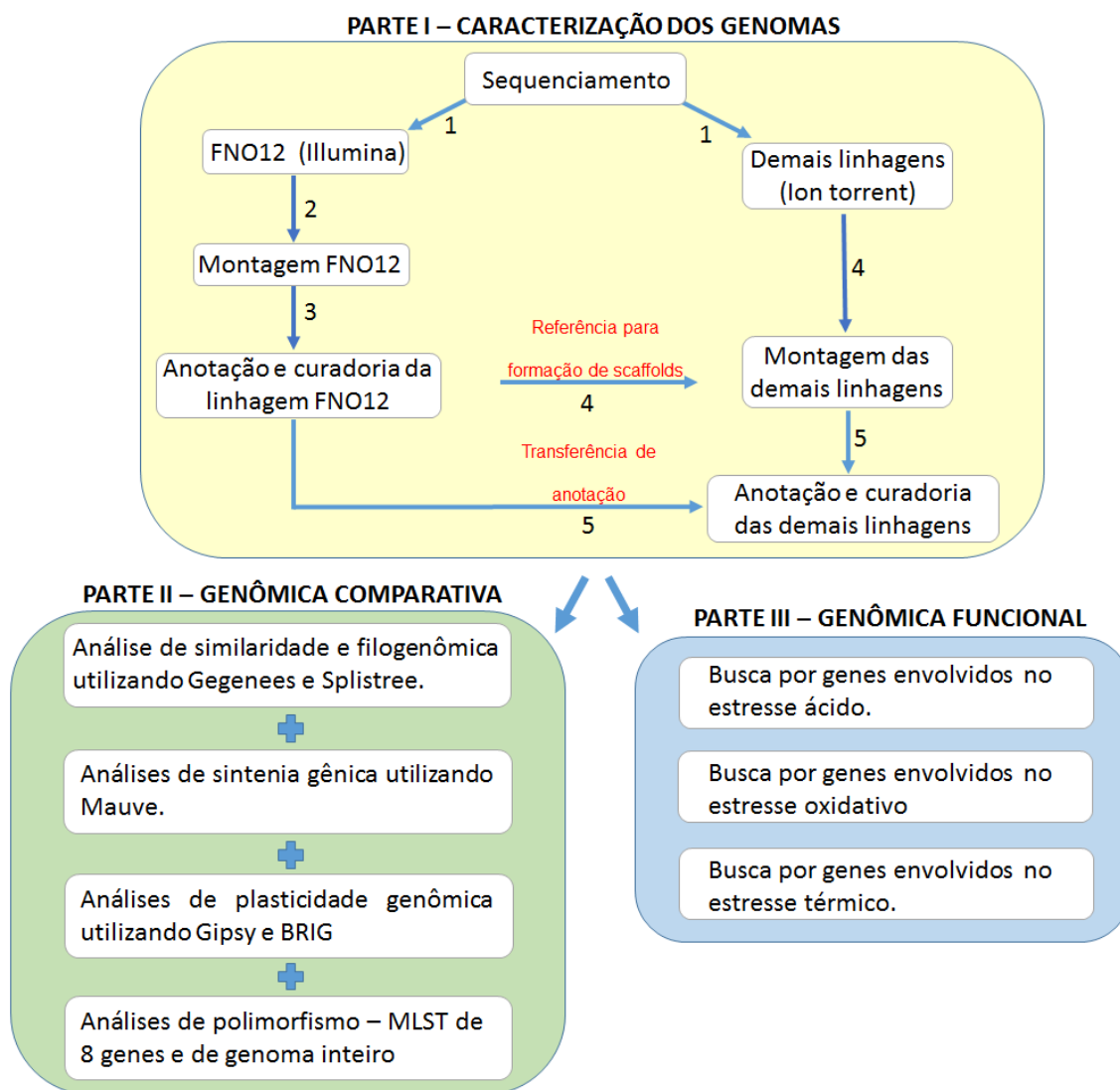


Figura 6: Desenho experimental das etapas para a caracterização das linhagens de FNO. Parte I indica os passos seguidos para a caracterização genômica das linhagens isoladas no Brasil. Os números próximos às setas azuis indicam a ordem seguida na execução dessa etapa. Parte II mostra as análises e os programas que foram utilizados para a comparação dos genomas. Parte III da caracterização das linhagens consistiu na busca por genes possivelmente envolvidos na resposta a estresses ambientais.

5.1. Obtenção do DNA genômico para sequenciamento

As linhagens de *Francisella noatunensis* subsp. *orientalis* FNO01, FNO12, FNO24, FNO39, FNO 44, FNO61, FNO75, FNO93, FNO95, FNO111, FNO117, FNO135, FNO190, FNO191, FNO205, FNO215 e FNO222 foram isoladas de surtos distintos que ocorreram em fazendas de piscicultura, sem conexão epidemiológica evidente, da região sudeste do Brasil (Tabela 4) (Figura 7).

Tabela 4: Linhagens de FNO que tiveram DNA genômico sequenciado; localização das fazendas de cultivo de tilápia que tiveram surto registrado e órgão de qual a respectiva linhagem foi isolada.

Strain	Município / País	Órgão
FNO01	Areado - MG	Rim
FNO12	Areado - MG	Rim
FNO24	Alterosa - MG	Baço
FNO39	Capim Branco - MG	Rim
FNO44	Alfenas - MG	Baço
FNO61	Ribeirão da Neves - MG	Rim
FNO75	Ribeirão da Neves - MG	Rim
FNO93	Fama - MG	Rim
FNO95	Passos - MG	Rim
FNO111	Varginha - MG	Rim
FNO117	Guapé - MG	Rim
FNO135	Alfenas - MG	Rim
FNO137	Santa Fé do Sul - SP	Rim
FNO190	Santa Fé do Sul - SP	Baço
FNO191	Três Marias - MG	Rim
FNO205	Carmo do Rio claro - MG	Rim
FNO215	Linhares - ES	Rim
FNO222	Linhares - ES	Rim

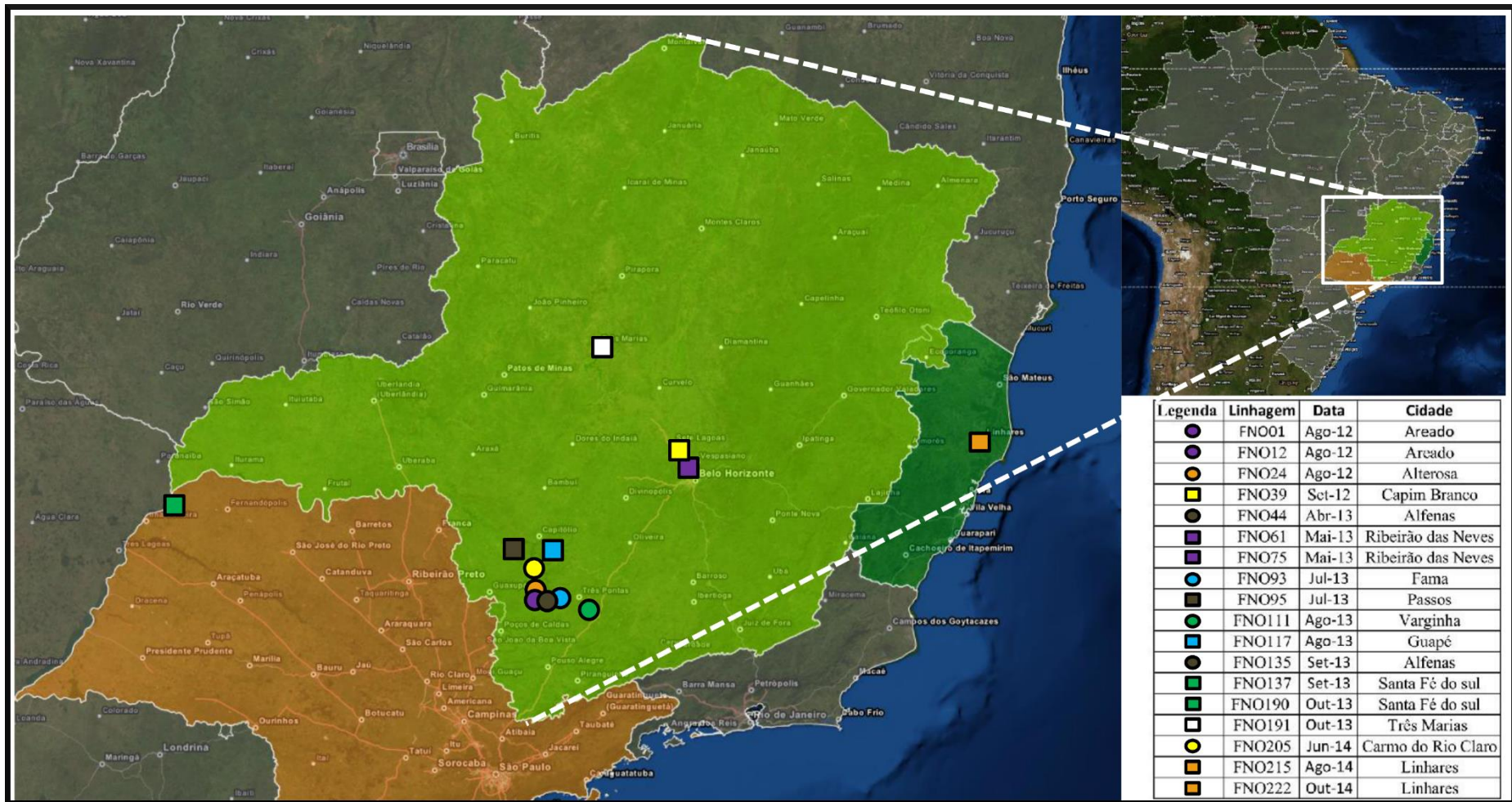


Figura 7: Localização geográfica de fazendas de cultivo de Tilápia do Nilo onde ocorreram surtos de franciselose e data de isolamento das linhagens.

Tais linhagens foram cultivadas em meio CHA (ágar cisteína coração com 2% de suplementação de hemoglobina bovina) a 28°C por um período de 96 horas. Para cada uma das linhagens foram realizados testes relativos à atividade de oxidase e catalase e testes primários de identificação bacteriana Gram. Para o armazenamento das amostras foi utilizado o caldo Mueller-Hinton cátion ajustado, suplementado com 1% de glicose, suplemento Vx (Laborclin, Brasil), além de 15% de glicerol.

Quando necessário essas amostras foram descongeladas e plaqueadas, novamente, no meio sólido citado anteriormente e incubadas a 28°C por 48 horas. As colônias que cresceram em placa foram retiradas com auxílio de uma alça descartável de 10ul e seu DNA genômico foi extraído utilizando o Maxwell® 16 Tissue DNA Purification Kit no aparelho *Maxwell 16® Research Instrument* (Promega, USA), de acordo com as recomendações do fabricante que consistem em quatro passos automatizados: i – lise das amostras na presença de um agente caotrópico e detergente; ii – ligação dos ácidos nucleicos a partículas de sílica magnetizadas; iii – lavagem das partículas ligadas, isolando-as dos outros componentes celulares; iv – eluição de ácidos nucleicos em tampão Tris-EDTA (do inglês *Ethylenediamine tetraacetic acid*). O DNA extraído foi acondicionado a -80°C até o uso.

5.2. Sequenciamento

5.2.1. Sequenciamento utilizando Ion Torrent PGM™

A etapa de construção das bibliotecas seguiu as recomendações do fabricante (Ion Xpress™ Plus gDNA Fragment Library Preparation, MAN0009847). Desse modo, a primeira etapa do processo de sequenciamento foi a fragmentação do DNA genômico utilizando-se o kit Ion Shear™ Plus Reagents. O tamanho dos fragmentos gerados por este reagente é tempo-dependente, sendo assim, seguiu-se um protocolo pré-estabelecido de 5 minutos a 37°C, para que o tamanho médio dos fragmentos de DNA fosse de 200pb.

A etapa posterior à fragmentação foi a ligação entre essas frações dos fragmentos de DNA à adaptadores, Ion Xpress™ Barcode Adapters, seguido de *nick-repair* que garante a ligação correta entre eles. Nesta fase foi feita a escolha da biblioteca baseada no tamanho do fragmento desejado – o que dá o nome a etapa de *Size-selection*. Em sequência foi feita a normalização das bibliotecas em aproximadamente 100pM utilizando o Ion Library Equalizer™ kit. O fluxo de trabalho com o kit consistiu em três passos: i – amplificação da biblioteca com primers; ii – captura da biblioteca com *beads* e iii - eluição em tampão. Após a fase supracitada,

seguiu-se a etapa de amplificação das bibliotecas seguindo as instruções do fabricante (Ion PGM™ Template OT2 200 Kit, MAN0007220). Nesse processo, os fragmentos foram amplificados por meio de uma PCR em emulsão onde as bibliotecas se ligaram a partículas esféricas (Ion Sphere™ particles) por meio do sistema Ion One Touch™ e dessa forma ocorreu uma amplificação clonal. As partículas ligadas a fragmentos de DNA são chamadas de positivas e separadas no Ion OneTouch™ ES, processo denominado enriquecimento.

Após a etapa de enriquecimento, às esferas positivas foram adicionados o Control Ion Sphere™ Particles, Sequencing Primer e Ion PGM™ Sequencing Polymerase (Ion PGM™ Sequencing 200 Kit). Em seguida, a reação foi carregada no chip semicondutor (Ion 318™ Chip v2) e inserido no Ion PGM™.

5.2.2. Sequenciamento utilizando Illumina® MiSeq

O sequenciador Illumina® MiSeq foi utilizado apenas no sequenciamento da linhagem de referência do trabalho, FNO12. Assim como na plataforma de sequenciamento descrita na seção anterior, o processo se iniciou com a fragmentação do DNA genômico. A estes novos fragmentos foram adicionados, em suas extremidades, adeninas que permitiram a ligação de adaptadores. Esses fragmentos foram, então, selecionados com base no peso molecular, purificados e fixados a um suporte sólido por meio da adesão dos adaptadores a oligonucleotídeos que são complementares a eles. Desse modo, foi feita a amplificação de cada um desses fragmentos dando origem a vários clusters de moléculas idênticas, todas ligadas ao suporte.

Em seguida, as fitas reversas foram clivadas, lavadas e primers para sequenciamento ligados a fita molde e, finalmente, todos os nucleotídeos marcados foram adicionados simultaneamente. A cada incorporação de nucleotídeo um sinal de luz era emitido e captado pelo detector.

5.3. Montagem

A primeira análise realizada foi com a ferramenta FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) para os dados obtidos no sequenciamento das duas plataformas, buscando identificar o tamanho e qualidade das sequências obtidas.

Após a confirmação que os dados de sequenciamento possuíam uma qualidade aceitável, ou seja, mais de 85% das reads com qualidade média de 20 na escala phred e com tamanhos esperados para cada tecnologia, iniciou-se uma busca por montadores

etapa, procurou-se atentar para que todas as contigs de todas as linhagens fossem aproveitadas, buscando assim identificar possíveis regiões distintas da referência. Também nesta etapa, foram identificadas quatro regiões que, sistematicamente, apresentavam baixa cobertura de reads. Para resolução de tais regiões, foram desenhados pares de primers flanqueadores e executado o sequenciamento capilar na plataforma ABI 3.500 (Applied Bioscience, EUA), inclusive para confirmação destas regiões na linhagem FNO12.

Além disso, apenas para que fosse possível a inclusão das linhagens *Francisella noatunensis* subsp. *orientalis* FNO770 e FNO771 e *Francisella noatunensis* subsp. *noatunensis* FNN178, FNN769, FNN772, FNN774, FNN775 e FNN846, em análises do presente estudo, foram feitos downloads dos dados brutos (*reads*) das mesmas no banco de dados do NCBI (<http://www.ncbi.nlm.nih.gov/bioproject/?term=francisella>) e realizada uma montagem automática utilizando o montador Mira para obtenção de contigs, sem fechamento manual de gaps.

5.4. Anotação

5.4.1. Predição e anotação automática dos genes

Foi realizada uma predição / anotação gênica utilizando o RAST (do inglês, Rapid Annotations using subsystems Technology – disponível no endereço eletrônico <http://rast.nmpdr.org/>). Trata-se de um serviço de anotação automática que pesquisa por similaridade protéica e indica prováveis genes em uma sequência nucleotídica sejam eles RNAs estruturais ou não. Para que o processo fosse feito, foi realizado um cadastro no site e submetido o genoma no formato fasta. Em seguida, o arquivo contendo a anotação automática foi submetido ao processo de curadoria manual de anotação e frameshifts.

5.4.2. Curadoria Manual

O programa Artemis (disponível no endereço eletrônico <http://www.sanger.ac.uk/resources/software/artemis/>) permite a visualização dos genomas e realização da anotação manual. No processo de curadoria manual o anotador pode verificar as posições dos códons iniciais e finais das CDSs (do inglês, Coding Sequence) e, quando necessário, trinar as mesmas para o início correto em uma metionina ou estendê-las até um códon de parada.

Além disso, é possível fazer pesquisas baseadas em similaridade nos bancos de dados biológicos como Uniprot (disponível no endereço eletrônico

http://www.uniprot.org/). Este contém informações anotadas manualmente (TrEMBL) e informações que além de anotadas foram revisadas (Swiss-Prot). Desse modo, as CDSs recebiam atributos como produto e gene (quando possível) conforme os resultados de BLAST (do inglês, Basic Local Alignment Search Tool) realizados no mesmo. Nesse processo foi dado uma importância maior (ou relevância maior) à informações de produtos revisados, o quanto determinado nome de produto apareceu para cada resultado, a identidade do alinhamento e o valor do e-value (Figura 9).

Por fim, CDSs que não apresentavam nenhum resultado de BLAST ou as que pesquisas pelo nome do produto eram inconclusivas foram consideradas como proteínas hipotéticas, assim como a maioria daquelas que não foram preditas pelo RAST e foram criadas pelo anotador.

The image displays a workflow for manual curation in Artemis. **A)** The 'Feature base viewer for feature:uhpT' window shows the amino acid sequence: ALFGAHVFDQVIGMIFFFSVLCLIGFGLRVGNDSPESVGLGPVEEIFDEPISSEDA... **B)** The UniProt BLAST search results show a top hit with an E-value of 0.0 and a score of 2405. **C)** The alignment window shows the query sequence aligned with the reference sequence. **D)** The 'Artemis Feature Edit: uhpT' window shows the updated annotations: /colour=3, /gene="uhpT", /product="Sugar phosphate antiporter", /note="Alinhamento de 100%", and /locus_tag="FNO191_1843".

Figura 9: Fluxo de trabalho da curadoria manual. **A)** Com auxílio do Artemis foi possível copiar a sequência de aminoácidos. **B)** Tal sequência é colada no banco de dados Uniprot para que seja feito o blast. **C)** Os resultados mostram o possível produto e gene a serem anotados, além disso, o alinhamento serve como referência para possíveis ajustes nas CDSs. **D)** Os resultados obtidos no Uniprot podem ser registrados na CDS em seus respectivos campos.

5.4.3. Curadoria de frameshifts

O processo de curadoria de pseudogenes baseia-se na análise de todos os frameshifts em busca de um possível erro gerado pelo processo químico de

sequenciamento. São averiguados um por vez com atenção especial para a região de mudança da matriz de leitura. Comumente, essas regiões de quebra são caracterizadas por homopolímeros e para que seja feita uma avaliação detida sobre a real sequência de nucleotídeos nessa região foi utilizado simultaneamente dois programas: Artemis e CLC Genomics Workbench. O último permite que seja feito um mapeamento das reads da montagem com o genoma que já foi finalizado. Assim, uma parte da CDS, geralmente a região da quebra gerada pelo frameshift, é selecionada no Artemis e finalizada no CLC. A partir da observação da cobertura e conservação das bases no último, a decisão por uma correção via deleção, inserção ou troca de bases é feita no Artemis (Figura 10). Finalmente, com o objetivo de garantir que aquelas CDSs que passaram por alguma edição estivessem posicionadas corretamente, foram feitos alinhamentos no BLAST do Uniprot.

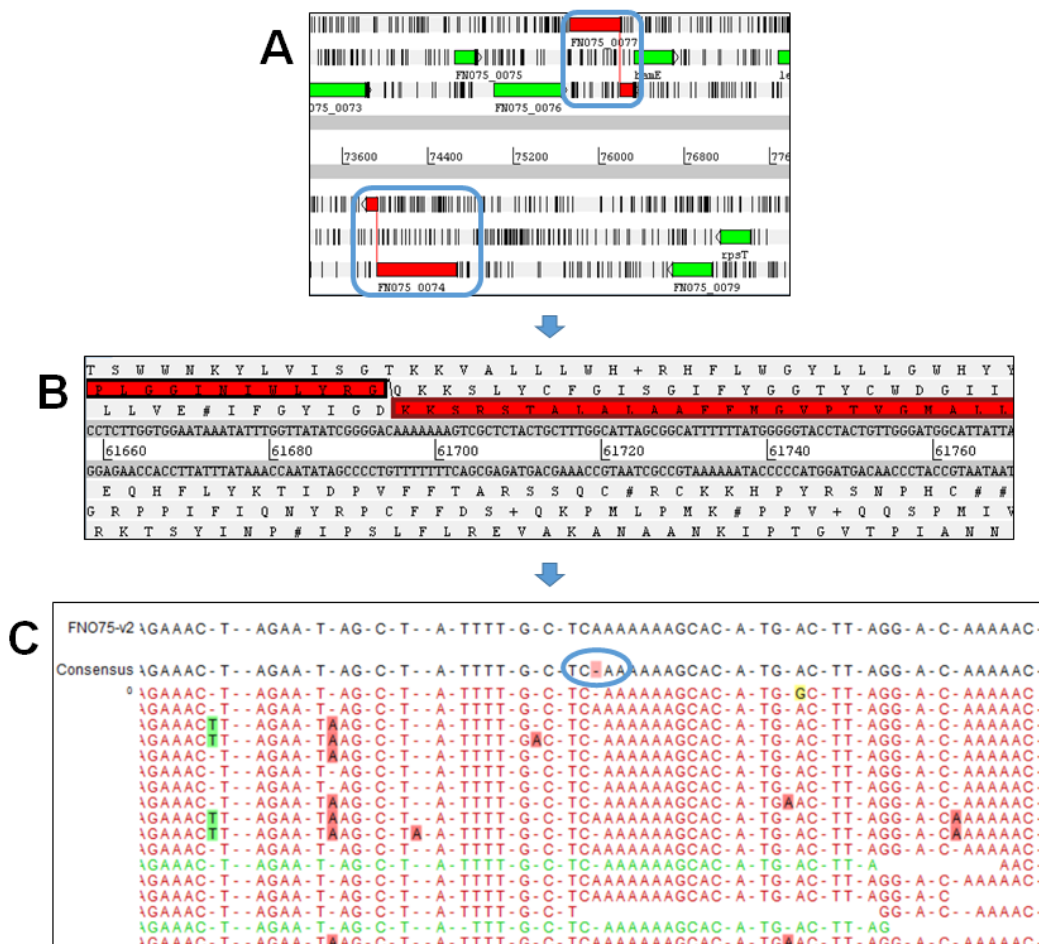


Figura 10: Fluxo de trabalho da curadoria de frameshifts. **A)** Com auxílio do Artemis as CDSs que apresentam uma quebra são selecionadas. **B)** Essas passam por uma avaliação da região de interrupção. **C)** Tal sequência é procurada no CLC Genomics para que, com base na cobertura, seja tomada a decisão de trocar, apagar, acrescentar ou manter alguma base. Qualquer alteração é feita no Artemis.

Após todo o processo de anotação e curadorias ser consolidado para a linhagem *Francisella noatunensis* subsp. *orientalis* FNO12, sequenciada com Illumina® Miseq, foi feita a transferência da anotação desta linhagem para todas as outras linhagens, que foram sequenciadas com Ion PGM™. Para tanto, foram utilizados *in-house scripts* e o arquivo de extensão .ffn que continha todas as sequências de nucleotídeos das CDSs e suas respectivas informações (posição, produto e gene).

Além disso, para cada genoma, foram realizadas predições de tRNA com o tRNAscan-SE Search Server (disponível no endereço eletrônico <http://lowelab.ucsc.edu/tRNAscan-SE/>) e predições de rRNA com o RNAmmer (disponível no endereço eletrônico <http://www.cbs.dtu.dk/services/RNAmmer/>). Nas duas situações, um arquivo de extensão fasta foi submetido e os *outputs* convertidos, com auxílio de *in-house scripts*, em arquivos com extensão .embl. Estes, por sua vez, podem ser abertos pelo software Artemis juntamente com o genoma que recebeu a transferência de anotação e, assim, salvos em um único arquivo com extensão .gbk (Figura 11).

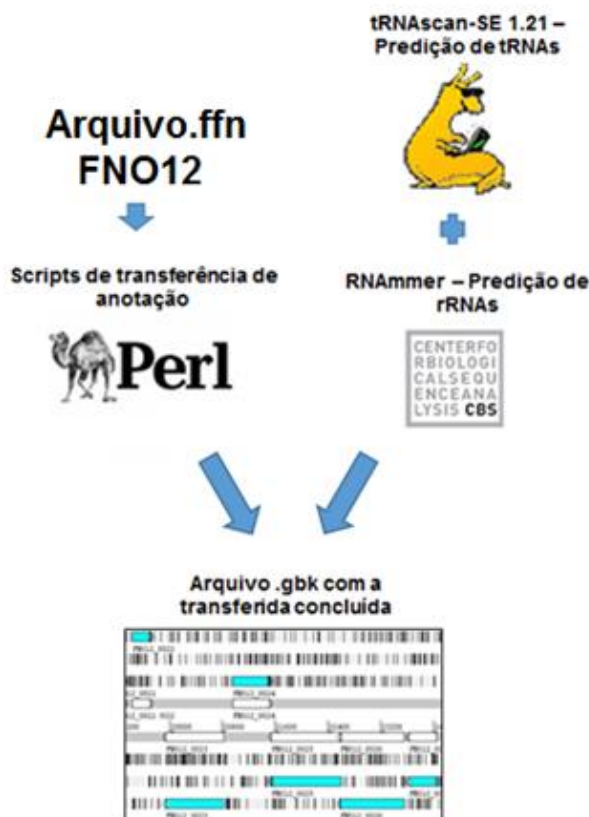


Figura 11: Fluxo de trabalho da transferência de anotação: com auxílio do Artemis foi gerado um arquivo de extensão .ffn contendo as informações de todas as CDS da linhagem FNO12. A transferência dessas informações para as outras linhagens é feita utilizando *in-house scripts*. Paralelamente, predições de RNAs estruturais são feitas nos sites do RNAmmer e tRNAscan-SE. Estes geram arquivos de extensão .embl que são convertidos, também por *in-house scripts*. Tanto o arquivo contendo a transferência de anotação e os arquivos com predições de RNAs são abertos no Artemis para que sejam salvos como um único arquivo.

5.5. Genômica comparativa

5.5.1. Plasticidade genômica

Com auxílio do software GIPSy (disponível no endereço eletrônico http://www.bioinformatics.org/groups/?group_id=1180) foram feitas predições de ilhas genômicas (GEI - do inglês, Genomic Island). Tal ferramenta possibilita a predição de quatro tipos de ilhas: patogenicidade, resistência, metabólica e simbiótica. Como já existe relatado na literatura uma ilha de patogenicidade para o gênero e, principalmente, como se trata de um patógeno optou-se pela predição de possíveis ilhas de patogenicidade para os organismos analisados no presente trabalho.

O primeiro e, possivelmente o mais importante passo dessa análise consiste na escolha de um organismo próximo filogeneticamente, preferencialmente não patogênico, para ser usado como referência durante o processo. Assim, a escolha da referência foi pautada em algumas premissas básicas: i. não ser patógeno; ii. ser o mais próximo possível filogeneticamente; iii. genoma completo e disponível.

Como se trata de um gênero de patógenos e praticamente todos os genomas sequenciados disponíveis da mesma ordem e classe são também de patógenos optou-se pela escolha de três genomas: *F. philomiragia* ATCC 25017, *F. tularensis* subsp. *novicida* U112 e *Thiomicrospira crunogena* XCL_2. Desse modo, os arquivos finais foram analisados e cruzados para que pudesse ser avaliado a coincidência das ilhas e, também, a força de predição uma vez que o software atribui pesos para as diferentes etapas do processo de análise.

Após a escolha do genoma de referência o software segue as seguintes etapas: i. análise de conteúdo G+C; ii. análises de uso de códon; iii. busca por genes de transposase; iv. busca por fatores específicos de virulência no banco de dados MvirDB (disponível no endereço eletrônico <http://mvirdb.llnl.gov/>); v. os genomas passaram pelo BLAST de proteínas (BLASTP), no qual todas as CDS do genoma de interesse foram comparadas contra todas as CDS do genoma de referência; vi. procura por genes de tRNA. Em seguida, o software agrupa todas as informações em um arquivo com as delimitações das ilhas. Tais análises foram realizadas utilizando FNO12.

5.5.2. Representação circular da comparação das sequências genômicas

O software BRIG (do inglês BLAST Ring Image Generator) foi utilizado para a visualização dos genomas tendo como referência a linhagem FNO12. Para tanto, foram necessários todos os genomas no formato fasta (.fna) para formação de um banco de dados e conseqüente realização de BLASTn. O programa também permitiu a escolha

da ordem dos genomas e inserção do arquivo tabular gerado pelo GIPSy. Assim, são criados anéis correspondentes as sequencias genômicas com as marcações das GEIs preditas.

5.5.3. Sintenia gênica

Foi, também, realizada uma análise de sintenia e para este fim foi utilizado o software de alinhamento múltiplo de genomas Mauve (disponível em <http://asap.ahabs.wisc.edu/mauve>). Os arquivos fasta contendo as sequências de nucleotídeos dos genomas completos serviram de “input”. O programa realiza alinhamentos múltiplos de todos os genomas e o parâmetro utilizado foi o padrão, recomendado pelo manual do usuário (disponível em <http://asap.genetics.wisc.edu/>) para organismos intimamente relacionados, *progressiveMauve*.

Após os alinhamentos, os genomas são dispostos em forma de blocos gênicos coloridos indicando homologia entre determinadas partes das sequencias de DNA genômico. Isto facilita a visualização e a comparação entre os mesmos.

5.5.4. Gegenees

Para que fosse feita a análise filogenética do gênero *Francisella* foi utilizado o software Gegenees (disponível em <http://www.gegenees.org/>). Desse modo, foram submetidos todos os genomas de interesse no formato .fna, entre eles a sequência de DNA genômico de *Dichelobacter nodosus* VCS1703A que serviu de *outgroup* e tinha como objetivo enraizar a árvore. Adicionalmente, os parâmetros utilizados foram default e utilizavam o método de BLASTN para fazer o alinhamento de todos os genomas, que foram fragmentados e tiveram todas as suas partes comparadas contra todas as outras dos genomas incluídos na análise. O resultado disso é um “heatplot” com valores que indicam o grau de similaridade entre eles em nível de nucleotídeos.

Tais valores foram salvos na extensão .nexus e serviram como arquivo de entrada para o software Splitstree (disponível em <http://www.splitstree.org/>), com o qual foi gerado uma árvore utilizando o método de *Neighbor-Joining* (NJ).

5.5.5. MLST de 8 genes

Com o intuito de obter uma descrição sistemática e padronizada de variação genotípica das linhagens de FNO sequenciadas, optou-se por uma abordagem gene a gene conhecida como MLST (do inglês *Multilocus sequence typing*). Para isso, foi utilizado o BIGSdb (do inglês *Bacterial Isolate Genome Sequence Database*), que é um

banco de dados de grande capacidade no qual muitas análises, como o MLST, podem ser feitas utilizando sequências genômicas provenientes de diferentes repositórios e tecnologias de sequenciamento.

Com o auxílio de um módulo específico desse banco de dados, *Genome Comparator*, 8 genes *housekeeping* (Tabela 5), comumente escolhidos para este tipo de estudo em bactérias, foram comparados por meio do uso de BLAST. Os genes de FNO12 foram escolhidos como referência e, assim, cada novo alelo foi contado como resultado de um único evento genético, independentemente do número de polimorfismos de nucleotídeos envolvidos (Maiden *et al.*, 2013).

Tabela 5: Genes *housekeeping* utilizados para análise de MLST das linhagens de *Francisella noatunensis* subsp. *orientalis*. As informações são referentes à FNO12.

Gene	Produto	Localização no genoma	Identificação
<i>murC</i>	UDP-N-acetylmuramate--L-alanine ligase	53396..54754	FNO12_0055
<i>dnaE</i>	DNA polymerase III subunit alpha	443763..447242	FNO12_0485
<i>pgm</i>	Phosphoglucomutase	472999..474633	FNO12_0504
<i>pheS</i>	Phenylalanine--tRNA ligase alpha subunit	957747..956734	FNO12_0983
<i>dnaG</i>	DNA primase	984671..982851	FNO12_1010
<i>tkt</i>	Transketolase	1330622..1328631	FNO12_1378
<i>sdhA</i>	Succinate dehydrogenase	1699085..1697292	FNO12_1747
<i>infB</i>	Translation initiation factor IF-2	1723141..1720607	FNO12_1771

O *Genome Comparator* produz dois *outputs*, o primeiro deles é uma tabela mostrando quais genes são idênticos à referência e quais são diferentes. O segundo trata-se de uma matriz de distância com base nos perfis alélicos identificados. Esta foi salva na extensão *.nexus* e usada como arquivo de entrada no software *Splitstree* para gerar uma árvore pelo método de *Neighbor-Joining*.

5.5.6. MLST de genoma inteiro ou abordagem gene a gene

O MLST de genoma inteiro (do inglês *whole-genome MLST – wgMLST*) foi feito com o auxílio do *BIGSdb*. Porém, todos os genes de todas as linhagens de FNOs foram comparados contra os genes da FNO12, escolhida como referência neste trabalho, acrescidos de duas CDSs que FNO24 possui a mais em relação as outras linhagens.

Desse modo, o procedimento para obtenção dos resultados é o mesmo detalhado na seção anterior, desde a utilização do *Genome Comparator* até a elaboração da árvore pelo método de *Neighbor-Joining*.

6. RESULTADOS E DISCUSSÃO

6.1. Caracterização dos genomas

6.1.1. Sequenciamento e Montagem

O processo de sequenciamento apresentou, para ambas as tecnologias, a característica de perda de qualidade nas últimas 20 bases das reads com tamanho acima das médias; que foram de 180bp para a plataforma Ion PGM™ e 280bp para o dataset produzido pelo Illumina® Miseq. Desse modo, a escolha da linhagem *Francisella noatunensis* subsp. *orientalis* FNO12 para ser a primeira no processo de montagens foi além de uma escolha temporal, pois ela foi uma das primeiras a ser sequenciada, uma decisão baseada nos dados gerados pelo sequenciamento como descrito na Tabela 5.

Comparando todas as linhagens que foram sequenciadas, a cobertura obtida para a linhagem FNO12 foi a maior, pouco mais de 1380 vezes. Além disso, com o auxílio do montador Edena foi obtido o terceiro menor número de contigs, 15, e um alto valor de N50, 275.043.

O processo de finalização da montagem contou com a ajuda de um mapeamento dos contigs com um mapa de restrição para formação de scaffold antes do processo de fechamento de gaps, o qual proporcionou uma confiabilidade maior à metodologia adotada. Por fim, para orientar o início do genoma de FNO12 foi escolhida a linhagem *Francisella noatunensis* subsp. *orientalis* Toba04 por ser um genoma completo da mesma espécie disponível no NCBI e por possuir uma alta similaridade de nucleotídeos.

Os procedimentos relativos às montagens das outras linhagens de FNO só foram iniciados após minuciosa curadoria manual da linhagem FNO12 e certeza de que se tratava da versão final do genoma. Isso se deve ao fato de que tal linhagem serviu como referência para as demais linhagens no processo de formação de scaffolds.

Dessa forma, os resultados, também resumidos na tabela 6, para todas as outras linhagens mostram que mesmo com uma profundidade de cobertura menor do que a alcançada para FNO12, os sequenciamentos das demais linhagens obtiveram bons resultados relativos a coberturas. Essas variaram de 79,82 vezes para FNO24 a 381,84 vezes para FNO01.

Além disso, por estratégia de trabalho do grupo, as primeiras linhagens foram montadas com o montador Mira, e posteriormente com o montador Newbler, que se mostrou mais apropriado e obteve melhores resultados. Tal diferença de desempenho também pode ser atribuído à capacidade deste último ser mais indicado pra trabalhar com dados com muitos erros. O primeiro montador obteve como resultados para maior contig valores entre 324.401 para FNO39 e 453.908 para FNO61; para números de contigs valores entre 22 para FNO39 e 65 para FNO61 e para N50 valores que variam

entre 87.100 para FNO24 e 325.113 para FNO44. Já o segundo montador obteve como resultados para maior contig valores entre 283.352 para FNO191 e 478.721 para FNO222; para números de contigs valores entre 13 para FNO205 e FNO222 e 35 contigs para FNO191 e, por fim para valores de N50 o menor foi o da linhagem FNO191, 153.726, e o maior para foi o da linhagem FNO205, 304.013.

Tabela 6: Resumo dos dados de sequenciamento e montagem das linhagens de *Francisella noatunensis* subsp. *orientalis* isoladas de surtos no Brasil.

Linhagem	Tecnologia	Kit	Biblioteca	Cobertura	Montador	Maior contig	Número de contigs	N50
FNO01	Ion torrent	Sequencing 200bp	Fragmento	381.54x	Newbler	478.696	14	303.981
FNO12	Illumina	Nextera v3	Fragmento	1382.15x	Edena	479.234	15	275.043
FNO24	Ion torrent	Sequencing 200bp	Fragmento	79.82x	Mira	360.755	57	87.100
FNO39	Ion torrent	Sequencing 200bp	Fragmento	164.47x	Mira	324.401	22	143.811
FNO44	Ion torrent	Sequencing 200bp	Fragmento	184.21x	Mira	325.113	25	325.113
FNO61	Ion torrent	Sequencing 200bp	Fragmento	112.48x	Mira	453.908	65	140.933
FNO75	Ion torrent	Sequencing 200bp	Fragmento	140.38x	Newbler	378.301	18	237.008
FNO93	Ion torrent	Sequencing 200bp	Fragmento	176.41x	Newbler	338.417	18	198.377
FNO95	Ion torrent	Sequencing 200bp	Fragmento	231.10x	Newbler	429.106	16	274.831
FNO111	Ion torrent	Sequencing 200bp	Fragmento	240.82x	Newbler	426.350	15	304.010
FNO117	Ion torrent	Sequencing 200bp	Fragmento	300.94x	Newbler	429.019	15	274.814
FNO135	Ion torrent	Sequencing 200bp	Fragmento	215.94x	Newbler	426.304	18	274.771
FNO137	Ion torrent	Sequencing 200bp	Fragmento	195.20x	Newbler	426.300	16	274.809
FNO190	Ion torrent	Sequencing 200bp	Fragmento	203.43x	Newbler	429.122	16	237.022
FNO191	Ion torrent	Sequencing 200bp	Fragmento	115.99x	Newbler	283.352	35	153.796
FNO205	Ion torrent	Sequencing 200bp	Fragmento	369.43x	Newbler	478.705	13	304.013
FNO215	Ion torrent	Sequencing 200bp	Fragmento	169.96x	Newbler	463.362	15	274.930
FNO222	Ion torrent	Sequencing 200bp	Fragmento	135.40x	Newbler	478.721	13	303.989

6.1.2. Anotação

Finalizadas as etapas de montagem, anotação e curadoria manual para todas as linhagens, foi possível observar que os genomas apresentaram um tamanho final que variou de 1.859.533 bp para FNO117 a 1.862.561 bp para FNO39 e um número de CDSs que ficou entre 1529 para FNO39 a 1545 para FNO117. Todas as linhagens apresentaram a mesma proporção de 32% de conteúdo G+C, a mesma quantidade de 50 RNAs estruturais para todas as linhagens, com a exceção da FNO24 que possui um tRNA a menos em relação às demais. Além disso, a quantidade de pseudogenes variou de 356 para FNO117 a 374 para FNO39 (Tabela 7). Por fim, em relação à predição gênica, foi observado uma média de 621 proteínas hipotéticas para todas as linhagens.

Desse modo, os dados obtidos para as linhagens brasileiras, foco deste estudo, evidenciam uma grande semelhança com as outras linhagens de FNO como a *Francisella noatunensis* subsp. *orientalis* Toba04, já caracterizada por Sridhar *et al.* (2012) e a *Francisella noatunensis* subsp. *orientalis* LADL-07-285A disponível no banco de dados do NCBI. A primeira linhagem apresenta um genoma de 1.847.202 pb, 1595 CDSs, 39 RNAs estruturais e 252 pseudogenes; já a segunda apresenta um genoma de 1.858.987 pb, 1624 CDSs, 50 RNAs estruturais e 317 pseudogenes.

Destacadamente as informações acerca do número de pseudogenes indicam que, assim como outras espécies do gênero, incluindo subespécies de *F. tularensis*, as linhagens brasileiras estariam passando por um processo de redução genômica (Larsson *et al.*, 2009; Kingry & Petersen, 2014). Além disso, durante a transição de um organismo de vida livre para um patógeno intracelular é comum que genes relacionados a vias metabólicas sejam eliminados, uma vez que os nutrientes podem ser adquiridos a partir do hospedeiro (Kingry & Petersen, 2014). Isso indica que FNOs podem estar passando pelo mesmo processo, ainda que os pseudogenes nos isolados sequenciados e anotados para este trabalho, foram em sua maioria, 42% em média, preditos como proteínas hipotéticas.

Por fim, a caracterização dos genomas de FNO, descritos nessa seção, contribuiu para a elaboração de um artigo que está em processo de submissão na *Standards in Genomic Sciences* (SIGS). O mesmo encontra-se na seção anexos.

Tabela 7: Resumo da caracterização dos genomas das linhagens de *Francisella noatunensis* subsp. *orientalis* isoladas em diferentes fazendas de cultivo no Brasil.

Linhagem	Tamanho do genoma (bp)	Conteúdo G+C (%)	CDS	Pseudogenes	rRNAs	tRNAs	tmRNAs
FNO01	1.859.830	32%	1545	356	10	39	1
FNO12	1.859.720	32%	1538	363	10	39	1
FNO24	1.862.322	32%	1537	366	10	38	1
FNO39	1.862.561	32%	1529	374	10	39	1
FNO44	1.859.758	32%	1535	366	10	39	1
FNO61	1.859.834	32%	1537	364	10	39	1
FNO75	1.859.595	32%	1533	368	10	39	1
FNO93	1.859.715	32%	1539	362	10	39	1
FNO95	1.859.726	32%	1539	362	10	39	1
FNO111	1.859.667	32%	1544	357	10	39	1
FNO117	1.859.533	32%	1545	356	10	39	1
FNO135	1.859.534	32%	1536	365	10	39	1
FNO137	1.859.600	32%	1542	359	10	39	1
FNO190	1.859.595	32%	1539	362	10	39	1
FNO191	1.859.595	32%	1539	362	10	39	1
FNO205	1.859.728	32%	1541	360	10	39	1
FNO215	1.859.821	32%	1542	359	10	39	1
FNO222	1.859.795	32%	1544	357	10	39	1

6.2. Genômica comparativa

6.2.1. Gegenees

Devido à alta semelhança entre as linhagens de FNO isoladas no Brasil, demonstrada nos dados da seção anterior, optou-se por realizar análises de genômica comparativa para estudar a hipótese de uma relação clonal entre as mesmas. Desse modo, foi feito, primeiramente, uma análise utilizando o Gegenees que gerou um *heatmap* baseado em resultados de BLAST que indicam o grau de similaridade entre os genomas em nível de nucleotídeos (Figura 12).

Nesse figura é possível observar que as cores variaram numa escala de verde a laranja, equivalentes às porcentagens de similaridade existentes nos fragmentos melhor alinhados. Esses valores são utilizados como uma matriz de distância para a criação de uma árvore filogenômica no Splitstree, representada a esquerda da mesma figura.

Os resultados apresentados mostram que é possível distinguir clusters formados basicamente por linhagens de cada subespécie: *F. tularensis* subsp. *holartica*, *F. tularensis* subsp. *tularensis*, *F. tularensis* subsp. *novicida*, *F. philomiragia* subsp. *philomiragia*, *Francisella* sp, *F. noatunensis* subsp. *noatunensis* e *F. noatunensis* subsp. *orientalis*. Houve uma única exceção, a linhagem *F. cf. novicida* 3523 (posição 3) mesmo sendo representada próxima de outras linhagens da mesma subespécie (posições 4 e 5) apresenta uma menor similaridade de nucleotídeos, 75%, em relação as mesmas. Enquanto as linhagens *Francisella tularensis* subsp. *novicida* Fx1 e U112 apresentam uma similaridade de 95% entre elas. Com o *heatmap* é possível observar que a similaridade é maior, 100%, quando se trata das linhagens de FNO (posições 26 a 47) e das subespécies de *Francisella tularensis* subsp. *holartica* (posições 13 a 17) e *Francisella tularensis* subsp. *tularensis* (posições 7 a 12). Apenas as linhagens FNO770 (posição 27) e FTT_WY96_3418 (posição 7) apresentam similaridade de 99% em relação as outras linhagens da suas próprias subespécies. Esses resultados são compatíveis com análises filogenéticas feitas por Colquhoun *et al.*, 2011; Ahlinder *et al.*, 2012; Sjödin *et al.*, 2012 e Sridhar *et al.*, 2012;

Ademais, nos resultados aqui apresentados foi possível observar uma divisão entre dois grupos, o primeiro formado por linhagens de FNO, FNN, *F. philomiragia* ATCC25017 e *Francisella* sp. TX077308 e o segundo formado por subespécies de *Francisella tularensis*. Essa distinção já havia sido descrita na literatura e identificada como “grupo *Philomiragia*” (predominantemente patógenos de peixe) e “grupo *Tularensis*” (predominantemente patógenos de humano) (Sridhar *et al.*, 2012).

Desse modo, pode-se dizer, novamente que as linhagens de *Francisella noatunensis* subsp. *orientalis* são muito semelhantes e não foram encontradas diferenças que tornassem possível a distinção entre as espécies isoladas no Brasil e as demais. Isso nos leva a crer que diferenças devem existir em outros níveis, como rearranjos cromossômicos e polimorfismos. Porém, mesmo que as diferenças entre esses isolados seja pequena, é interessante destacar que em nível de nucleotídeos FNOToba04 é mais próxima das linhagens encontradas no Brasil do que FNOLADL-07-285A.

6.2.2. Sintenia gênica

A despeito do alto grau de similaridade em nível de nucleotídeos entre as linhagens de FNO observado nos resultados do obtidos com o Gegenees, uma característica marcante entre os genomas sequenciados de *Francisella* é o alto grau de rearranjos genômicos (Rohmer *et al.*, 2007). Desse modo, foi feita análise de sintenia gênica entre os genomas e para tanto foi utilizado o software Mauve, o qual é uma ferramenta para visualização da sintenia existente entre várias sequências genômicas de interesse. A comparação é baseada na homologia entre os genomas, objetivando dispor a sintenia através de blocos gênicos ajudando, dessa forma, na identificação de possíveis rearranjos entre os genomas (Darling *et al.*, 2010).

Comparando as linhagens de *Francisella noatunensis* subsp. *orientalis* FNO01, FNO12, FNO24, FNO39, FNO 44, FNO61, FNO75, FNO93, FNO95, FNO111, FNO117, FNO135, FNO190, FNO191, FNO205, FNO215, FNO222, FNO770, FNO771, FNOToba04 e FNOLADL-07-285A foi obtida a figura 13, na qual os genomas são mostrados na horizontal e os segmentos que são homólogos entre os genomas são representados por blocos coloridos da mesma cor. Quando existe uma inversão dos mesmos em relação ao genoma de referência, eles são deslocados para baixo da linha horizontal. Adicionalmente, a similaridade de nucleotídeos pode ser vista pelo tanto que o bloco é preenchido, assim como, a ausência (de alguma parte em relação ao genoma de referência, nesse caso a linhagem FNO01) ou presença de espaços entre os blocos significam, respectivamente, deleções e inserções.

Os resultados apresentados para essa metodologia mostram que as linhagens brasileiras, que ocupam as 18 primeiras posições da figura 10, possuem um alto grau de sintenia entre elas e é possível observar que os blocos gênicos apresentam o mesmo tamanho, mesma ordem e posições correlatas. Quando são levadas em consideração todas as linhagens, também é possível observar uma alta semelhança entre os genomas com grandes segmentos homólogos e alta similaridade de nucleotídeos. A FNOLADL-07-285A apresenta apenas uma transposição de diferença em relação as linhagens brasileiras e isso se deve provavelmente a montagem deste genoma, uma vez que ao consultarmos o arquivo gbk depositado no NCBI podemos notar que o gene *dnaA* não está logo no começo do genoma. Já a linhagem FNOToba04 apresenta inversões em relação às linhagens sequenciadas para este trabalho apesar de também possuir grandes blocos homólogos. Por fim, as linhagens FNO770 e FNO771 apresentam alto grau de similaridade com grandes blocos gênicos na mesma ordem, com a exceção de uma pequena transposição no início do genoma da FNO770 em relação as linhagens isoladas no Brasil.



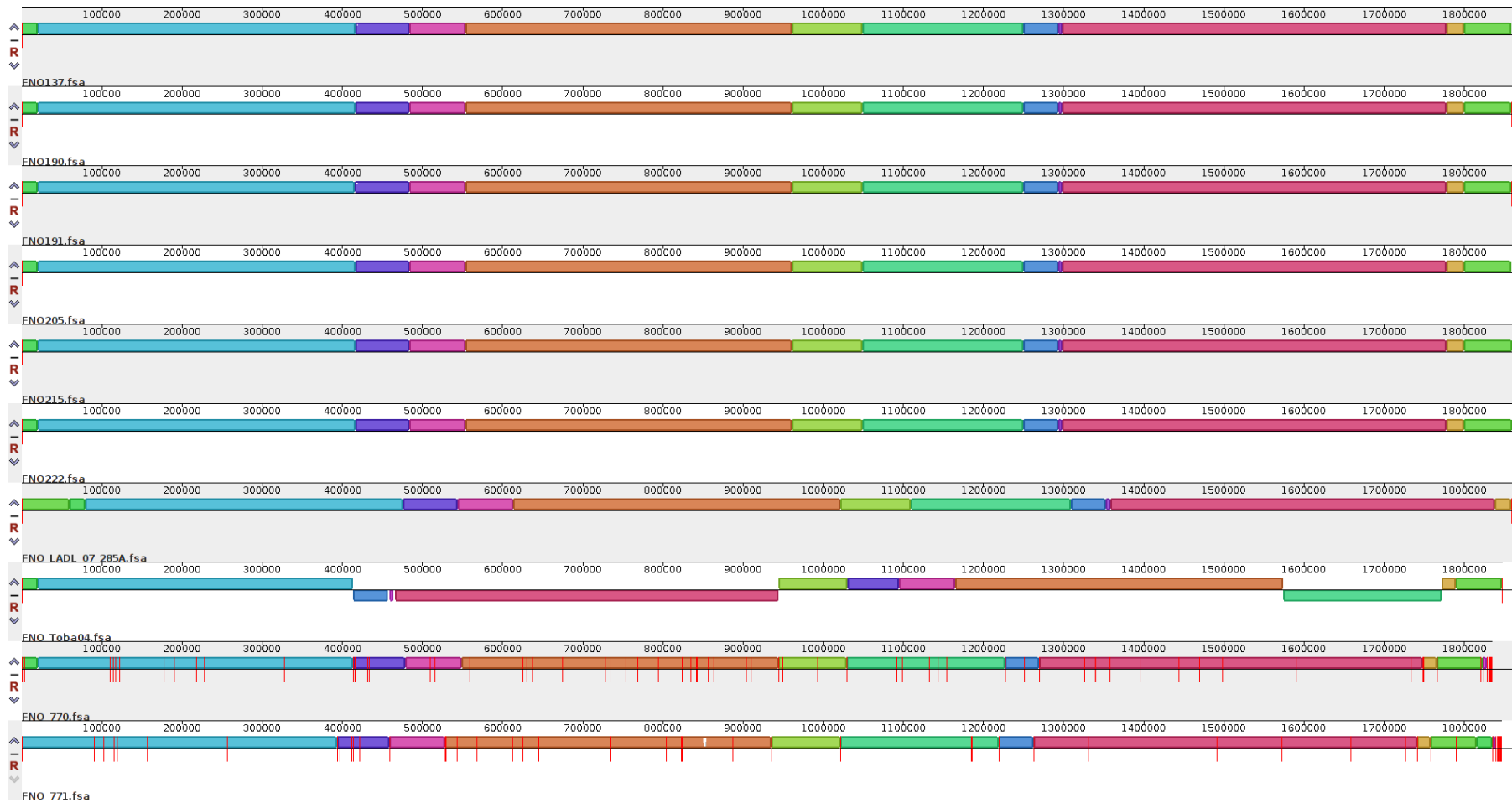


Figura 13: Análise de sintenia das linhagens de *Francisella noatunensis* subsp. *orientalis*. Os genomas são mostrados na horizontal e os segmentos que são homólogos entre os genomas são representados por blocos coloridos da mesma cor. Quando existe uma inversão dos mesmos em relação ao genoma de referência, eles são deslocados para baixo da linha horizontal.

Portanto, os resultados de sintonia corroboram a análise filogenômica vista anteriormente que mostram uma alta semelhança entre as linhagens, principalmente as linhagens isoladas no país. Além disso, evidenciam que o alto grau de rearranjos relatados por Rohmer et al. (2007) para o gênero não foi observado nessa comparação que contou apenas com linhagens de *Francisella noatunensis* subsp. *orientalis*.

6.2.3. Plasticidade genômica

A capacidade das bactérias de adaptação a diferentes ambientes resulta principalmente da sua elevada plasticidade genômica que é adquirida por meio de mutações pontuais, rearranjos e transferência de material genético (Dobrindt & Hacker, 2001; Barbosa *et al.*, 2014). A transferência horizontal de genes é o mecanismo necessário para a constituição de ilhas genômicas (GEIS). Essas, comumente, apresentam assinatura genômica do organismo de origem e isso pode ser constatado pelo diferente teor de conteúdo G + C e desvio no uso de códon em relação ao genoma do organismo receptor. Desse modo, GEIS podem conferir a uma célula novos e úteis fenótipos, tais como a capacidade de sobreviver dentro de um hospedeiro (Soares *et al.*, 2012).

Neste contexto, para que fosse feita a predição de possíveis eventos de transferência horizontal de genes nos genomas de *Francisella noatunensis* subsp. *orientalis* foi utilizado o GIPSy. Esta ferramenta é uma versão atualizada das metodologias utilizadas, até então, no PIPS (Pathogenicity Island Prediction Software) (Soares et al., 2012). Além de prever possíveis PAIs, a sua nova versão permite identificar ilhas simbióticas (SI), ilhas de metabolismo (MI) e ilhas de resistência (RI).

Para visualização desses resultados foi utilizado o software BRIG, no qual é possível observar anéis que representam os genomas e a similaridade de nucleotídeos existente entre as sequências genômicas e a referência. Essa similaridade pode ser notada pela intensidade das cores dos anéis, que varia de acordo com a identidade obtida por BLAST. Dessa maneira, foi obtida a figura 14 partes A e B.

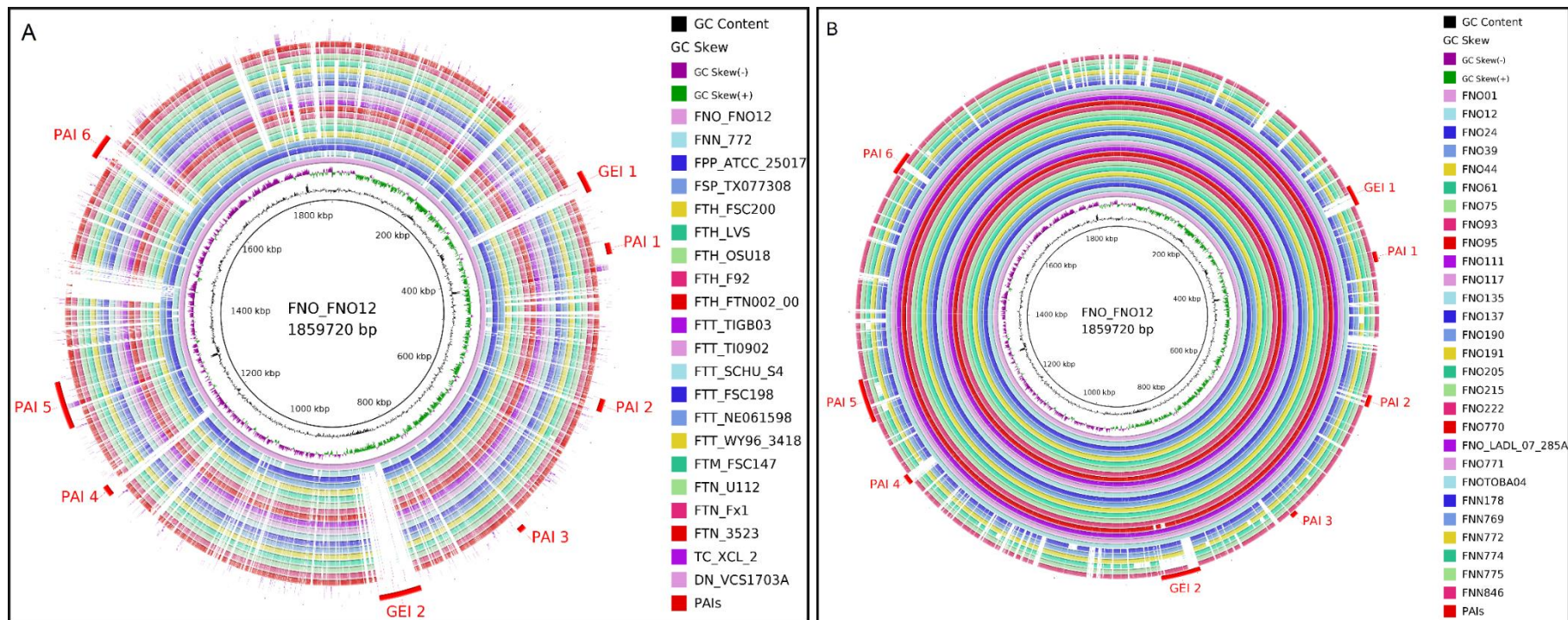


Figura 14: Análise de plasticidade genômica de *Francisella noatunensis* subsp. *orientalis*. As figuras **A** e **B** mostram GEIs e PAIs identificadas na linhagem FNO12 (referência). A intensidade das cores dos anéis na figura **A** e **B** mostra a similaridade de cada genoma em relação a referência, além de mostrar a presença ou ausência das GEIs e PAIs previstas para FNO12. Na seção **A** dessa figura pode-se distinguir o primeiro anel, que é a própria referência (FNO12), e em sequência as linhagens de *Francisella noatunensis* subsp. *noatunensis* (FNN), *Francisella philomiragia* subsp. *philomiragia* (FPP), *Francisella* sp. (FSP), *Francisella tularensis* subsp. *holartica* (FTH), *Francisella tularensis* subsp. *tularensis* (FTT), *Francisella tularensis* subsp. *mediasiatica* (FTM), *Francisella tularensis* subsp. *novicida* (FTN), *Thiomicrospira crunogena* (TC) e *Dichelobacter nodosus* (DN). Já a seção **B**, mostra as linhagens de *Francisella noatunensis* subsp. *orientalis* (FNO) e *Francisella noatunensis* subsp. *noatunensis* (FNN).

Baseado na intensidade das cores nota-se na parte A da figura 11, com exceção das linhagens *Francisella noatunensis* subsp. *noatunensis* FNN772 e *Francisella philomiragia* subsp. *philomiragia* FPP ATCC 25017, que a similaridade de nucleotídeos, em relação a referência não é alta. Percebe-se também que a semelhança é praticamente inexistente em relação a *Thiomicrospira crunogena* e *Dichelobacter nodosus*. Já a parte B evidencia que a similaridade de nucleotídeos das subespécies de *Francisella noatunensis* em relação a FNO12 é relativamente alta. Tal similaridade apresenta resultados mais elevados quando são consideradas somente as linhagens de FNO. Esses resultados corroboram o que já foi apresentado no presente trabalho com as análises feitas no Gegenees e Mauve.

Para *Francisella noatunensis* subsp. *orientalis* FNO12 foram preditas 8 ilhas genômicas, sendo 6 delas ilhas de patogenicidade (Figura 14). As PAIs 1 e 3 são observadas em todos os genomas e essa situação é semelhante para as PAIs 2 e 5 que também estão presentes em todos os genomas, mas em sua totalidade somente nas linhagens de FNO. A GEI 2 aparece quase que exclusivamente para as linhagens de FNO, pois apenas as linhagens FNN178, FNN769, FNN772 e FNN846 compartilham algumas partes dessa ilha. Nota-se também que a PAI 6 ocorre em sua totalidade para todas as linhagens de FNO, FPP ATCC 25017, FSP TX077308 e parcialmente para as demais linhagens. Por fim, também foram preditas a GEI 1 e a PAI 4 que seriam exclusivas das linhagens de FNO. Além disso, é importante destacar que nenhuma das ilhas citadas foram preditas para *Thiomicrospira crunogena* e *Dichelobacter nodosus*.

A busca por genes que constituem as ilhas preditas mostrou que a PAI 5 é composta por genes pertencentes à FPI (do inglês, *Francisella* pathogenicity island) (Tabela 8), que parecem codificar proteínas que compõem o sistema de secreção tipo VI (Hare & Hueffer, 2014). Tais proteínas são necessárias para o crescimento intracelular das espécies do gênero (Bruin 2011).

Tabela 8: Genes e produtos que compõem as ilhas genômicas preditas para FNO12

GEI / PAI	FNO12	Produtos
GEI 1	<i>rimO</i>	Ribosomal protein S12 methylthiotransferase
	<i>phrB</i>	Deoxyribodipyrimidine photolyase
	<i>thiE</i>	Thiamine-phosphate synthase
	<i>thiM</i>	Hydroxyethylthiazole kinase
	<i>thiL</i>	Thiamine-monophosphate kinase

	<i>thiD</i>	Phosphomethylpyrimidine kinase
PAI 1	<i>maeA</i>	NAD-dependent malic enzyme
	<i>Der</i>	GTPase Der
	<i>murA</i>	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
	<i>murB</i>	UDP-N-acetylenolpyruvoylglucosamine reductase
PAI2	<i>trxB</i>	Thioredoxin reductase
	<i>Gph</i>	Phosphoglycolate phosphatase
	<i>cutC</i>	Copper homeostasis protein
PAI3	<i>rpoH</i>	RNA polymerase sigma factor
	<i>aroA</i>	3-phosphoshikimate 1-carboxyvinyltransferase
	<i>rnhA</i>	Ribonuclease H
	<i>Ansa</i>	L-asparaginase 1
GEI 2	<i>icmF</i>	Type VI secretion protein IcmF
	<i>icmF2</i>	IcmF-related protein
PAI4	<i>neuA</i>	Acetylneuraminate cytidyltransferase
	<i>pseC</i>	UDP-4-keto-6-deoxy-N-acetylglucosamine 4-aminotransferase
	<i>pseB</i>	UDP-GlcNAc-specific C4,6 dehydratase/C5 epimerase
PAI 5	<i>mltA</i>	Membrane-bound lytic murein transglycosylase
	<i>rpIM</i>	50S ribosomal protein L13
	<i>rpsI</i>	30S ribosomal protein S9
	<i>mgIA</i>	Macrophage growth locus A
	<i>mgIB</i>	Macrophage growth locus B
	<i>rnhB</i>	Ribonuclease HII
	<i>yfhQ</i>	Methyltransferase
	<i>ptsN</i>	PEP-dependent sugar phosphotransferase
	<i>mnmE</i>	tRNA modification GTPase

	<i>cysK</i>	Cysteine synthase
	<i>pdpA</i>	Pathogenicity determinant protein
	<i>pdpB</i>	Pathogenicity determinant protein
	<i>iglD</i>	Intracellular growth locus protein D
	<i>iglC</i>	Intracellular growth locus protein C
	<i>iglB</i>	Intracellular growth locus protein B
	<i>iglA</i>	Intracellular growth locus protein A
	<i>pdpD</i>	Pathogenicity determinant protein D
	<i>anmK</i>	Anhydro-N-acetylmuramic acid kinase
PAI 6	<i>groL</i>	60 kDa chaperonin
	<i>groS</i>	10 kDa chaperonin
	<i>feoB</i>	Ferrous iron transport protein B
	<i>acpP2</i>	Acyl carrier protein

Comparando os genes que integram a FPI entre as linhagens *Francisella noatunensis* subsp. *orientalis* FNO12 e *Francisella tularensis* subsp. *novicida* U112, utilizada por Bruin *et al.* 2011 para o estudo desses genes, pode-se notar que os genes *pdpC* e *pdpE* não estão presentes nos patógenos de peixe. Segundo Sridhar *et al.* (2012), *pdpC* é, provavelmente, essencial para a infecção em células de mamíferos. Além disso, Broms *et al.* (2011) demonstraram que mutantes deficientes para o gene *pdpE* de *Francisella tularensis* subsp. *SCHU S4* tiveram fenótipos idênticos a linhagem parental em todos os testes realizados (replicação intracelular, escape fagossomal e atenuação em modelo murino). Portanto, pode-se sugerir também que o gene *pdpE* pode não ser essencial para infecção em peixes.

A deleção dos genes *iglA*, *iglB*, *iglC*, *iglD*, *pdpA*, *pdpB*, *pdpD*, *dotU* e *vgrG* mostrou que a replicação intracelular dos organismos do gênero depende dos mesmos (Bruin *et al.*, 2007; Bonquist *et al.* 2008; Ludu *et al.*, 2008; Broms *et al.*, 2012; Jia *et al.*, 2013; Robertson *et al.*, 2013 e Hare & Hueffer, 2014). Além disso, *iglC*, *iglD*, *iglE*, *iglG*, *iglI*, *dotU*, *pdpC* e *vgrG* estão relacionados ao escape fagossomal e, conseqüentemente,

relacionados à virulência destes organismos (Broms *et al.*, 2011; Bruin *et al.*, 2011; Broms *et al.*, 2012; Jia *et al.*, 2013; Lindgreen *et al.*, 2013; Robertson *et al.*, 2013).

De modo geral, os mecanismos moleculares que contribuem para a sobrevivência intracelular da *Francisella* são pouco compreendidos, e abordagens de mutagênese na FPI são úteis na identificação de genes necessários para a replicação intracelular e virulência (Hare & Hueffer, 2014). As informações disponíveis na literatura até o momento apenas sugerem as funções de genes como *mglA* e *mglB* como importantes para o crescimento intracelular de *Francisella*; a participação de *mltA* no processo de divisão celular; *rplM* e *rpsI* como genes que codificam proteínas ribossômicas em *Escherichia coli*; *rnhB* como gene relacionado a endonucleases que degradam RNA em *E. coli*; *mnmE* como gene envolvido na modificação de tRNAs também em *Escherichia coli* e *cysK* como gene relacionado a conversão de O-acetilserina para cisteína e sensibilidade para disponibilidade desse aminoácido em *Bacillus subtilis* (Baron & Nano, 1998; Lommatzsch *et al.* 1997; Isono *et al.* 1985; Itaya M, 1990; Cabedo *et al.* 1999; Tanous *et a.* 2008).

A GEI 1 apresentou os genes preditos *rimO*, *phrB*, *thiE*, *thiM*, *thiL* e *thiD*. O primeiro, de acordo com Anton *et al.* (2008), catalisa a metilação de proteínas ribossômicas, já o gene *phrB* está relacionado a reparação de danos causados por radiação no DNA (Kim & Sundin, 2001). Completando a ilha *thiE*, *thiM*, *thiL* e *thiD* formam um operon envolvido na biossíntese de tiamina (Reddick *et al.*, 2001; Mizote & Nakayama, 1989; Nishino, 1972; Petersen & Downs, 1997). Já a GEI 2 apresentou os genes *icmF* e *icmF2* e ambos estão associados à codificação de proteínas conservadas da membrana interna envolvidas no sistema de secreção tipo VI (Ma *et al.*, 2012).

A PAI 1 apresenta os genes *der*, *murA* e *murB*. Segundo Tan *et al.* (2002), o gene *der* codifica proteína que desempenha papel essencial nas etapas tardias da biogênese de ribossomos, uma vez que ela estaria envolvida na maturação de subunidades de ribossomo. Os genes *murA* e *murB* estão envolvidos na formação da parede celular (Marquardt *et al.*, 1992).

Dois genes podem ser destacados na PAI2, *gph* e *cutC*. O primeiro está envolvido no reparo de DNA (Kuznetsova *et al.*, 2006), já o gene *cutC* foi descrito como um dos genes que codificam proteínas que promovem homeostase de cobre em *Escherichia coli* (Li *et al.*, 2010).

Entre os genes que compõem a PAI 3 estão os genes *rpoH* e *aroA*. O primeiro codifica o fator sigma RpoH que está envolvido na resposta ao estresse térmico (Grall *et al.*, 2009) e o segundo tem como produto a sexta enzima da via chiquimato, que leva à síntese de compostos aromáticos. Como é uma via essencial para o crescimento de

microrganismos, o gene torna-se um alvo para o desenvolvimento de drogas (Eschenburg *et al.*, 2003).

Por fim, na PAI 6 podemos destacar os genes *groL*, *groS*, *feoB* e *acP2*, sendo que os dois primeiros estão envolvidos no dobramento e estabilização de proteínas em condições de estresse (Grall *et al.*, 2009). O gene *feoB* codifica proteína envolvida na absorção e transporte de ferro (Velayudhan *et al.*, 2000). Já o gene *acP2* está relacionado a transporte de acilos e síntese de ácidos graxos (Byers & Gong, 2007).

6.2.4. MLST e MLST de genoma inteiro ou abordagem gene a gene

A análise de abordagem gene a gene (MLST) foi proposta como um método baseado em sequências para identificar relações clonais entre bactérias (Maiden *et al.*, 2013) e desde então tem sido a força motriz no campo da epidemiologia molecular de patógenos. Uma característica importante que contribuiu para o sucesso foi a possibilidade de inferir padrões de descendência filogenética através da comparação dos perfis alélicos. Isso fez com que o MLST se tornasse o padrão-ouro para a vigilância epidemiológica de diversas espécies.

A variação alélica *housekeeping* é uma estratégia que tem sido amplamente explorada para a investigação de relações taxonômicas, incluindo a definição e identificação de espécies, e constitui a base de muitos estudos (Maiden *et al.*, 2013). Desse modo, a tipagem molecular é fundamental para determinar a estrutura da população e entender a evolução de bactérias patogênicas (Parkhill & Wren, 2011).

Determinar a presença de um surto, e se é clonal ou não, pode ajudar a direcionar as respostas para interrompê-lo, bem como para evitar surtos futuros e em um cenário ideal, onde são relatados todos os casos conhecidos, determinar a origem de um surto é trivial (Grad & Lipsitch, 2014).

Nesse contexto, a análise de abordagem gene a gene baseada em 8 genes *housekeeping* feita para as linhagens de *Francisella noatunensis* subsp. *orientalis* e *Francisella noatunensis* subsp. *noatunensis* não apresentou uma variabilidade capaz de determinar uma diferenciação entre as linhagens de FNO, principalmente se forem consideradas apenas as linhagens brasileiras. A linhagem FNOToba04 representou a única exceção, uma vez que ela apareceu destacada das demais linhagens de FNO (Figura 15). No entanto, pode-se dizer que a análise foi eficiente na discriminação em nível de subespécies, pois é possível observar dois grupos distintos: um formado por patógenos de bacalhau (FNN) e outro formado por patógenos de tilápia (FNO).

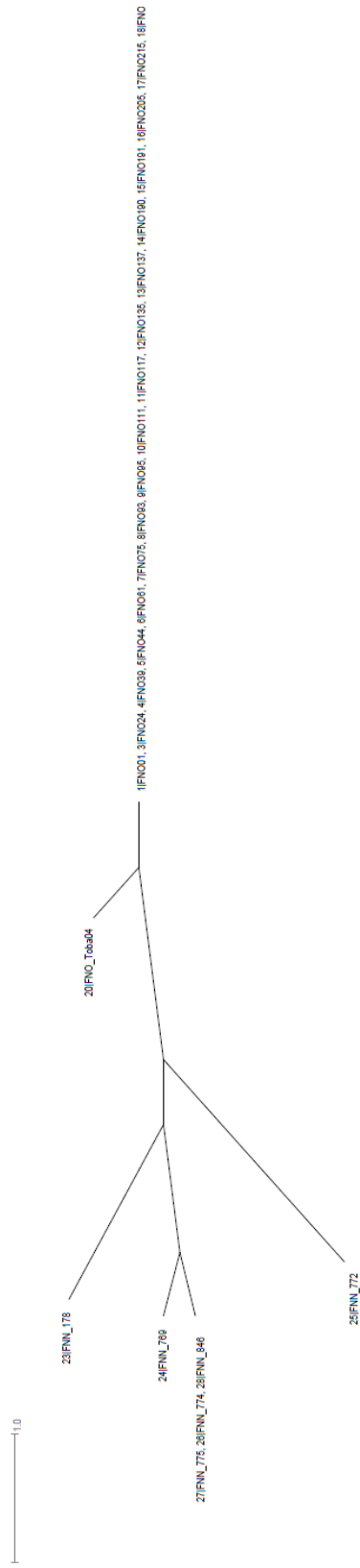


Figura 15: Relações filogenéticas das linhagens de *Francisella noatunensis* subsp. *orientalis* e *Francisella noatunensis* subsp. *noatunensis*. Resultado obtido por meio de análise de abordagem gene a gene baseada em genes *housekeeping*. O resultado indica a variabilidade dessas com em relação a esse perfil gênico.

Embora tal técnica seja altamente informativa, ela tem resolução limitada quando aplicada a isolados estreitamente relacionados. Assim, ela é muitas vezes inadequada para identificação de eventos evolutivos refinados ou para distinguir possíveis linhagens clonais dentro de uma epidemia recente (Parkhill & Wren, 2011).

Com a utilização de um número determinado de genes, como foi acima, pode-se obter uma de tipagem de resolução mais baixa, mas que seja suficiente para determinar as espécies que causam uma infecção, enquanto que níveis mais elevados de resolução podem ser alcançados mediante o aumento do número de genes incluídos na análise (Maiden *et al.*, 2013). A maior vantagem de utilizar genomas inteiros é o poder de discriminar entre linhagens, permitindo a geração de filogenias robustas (Parkhill & Wren, 2011). Em consequência de tais constatações, optou-se pela extensão da análise para que todos os alelos fossem considerados, técnica conhecida como MLST de genoma inteiro.

Para a análise de genoma inteiro, foram consideradas apenas as linhagens de *Francisella noatunensis* subsp. *orientalis* afim de analisar a origem e a evolução dessas, principalmente, das linhagens brasileiras. Desse modo, foi obtida a Figura 16, na qual é possível observar uma separação das linhagens isoladas no Brasil (FNO01 a FNO222) em relação às demais linhagens que foram encontradas no Japão (FNO771), Costa Rica (FNOLADL-07-285A e FNO770) e Indonésia (FNOToba04). É interessante notar, também, que, FNOToba04 é a linhagem mais distante das que aqui foram sequenciadas e anotadas, o que aponta para uma maior diferença na quantidade de polimorfismos entre as mesmas.

Além disso, pode-se sugerir, que a entrada do agente etiológico dos surtos de franciselose no Brasil ocorreu por entrada única de uma linhagem da espécie que foi se disseminando por diversas fazendas enquanto sofria uma diversificação em seu conteúdo gênico, através do tempo. Isso pode ser notado pela posição central da linhagem FNO12 na árvore filogenética, isolada do primeiro surto relatado no Brasil, em relação às demais linhagens brasileiras.

Por fim, as análises de polimorfismos somadas aos demais resultados obtidos no presente trabalho indicam uma alta clonalidade das linhagens de FNO, principalmente as linhagens isoladas de surtos relatados no Brasil.

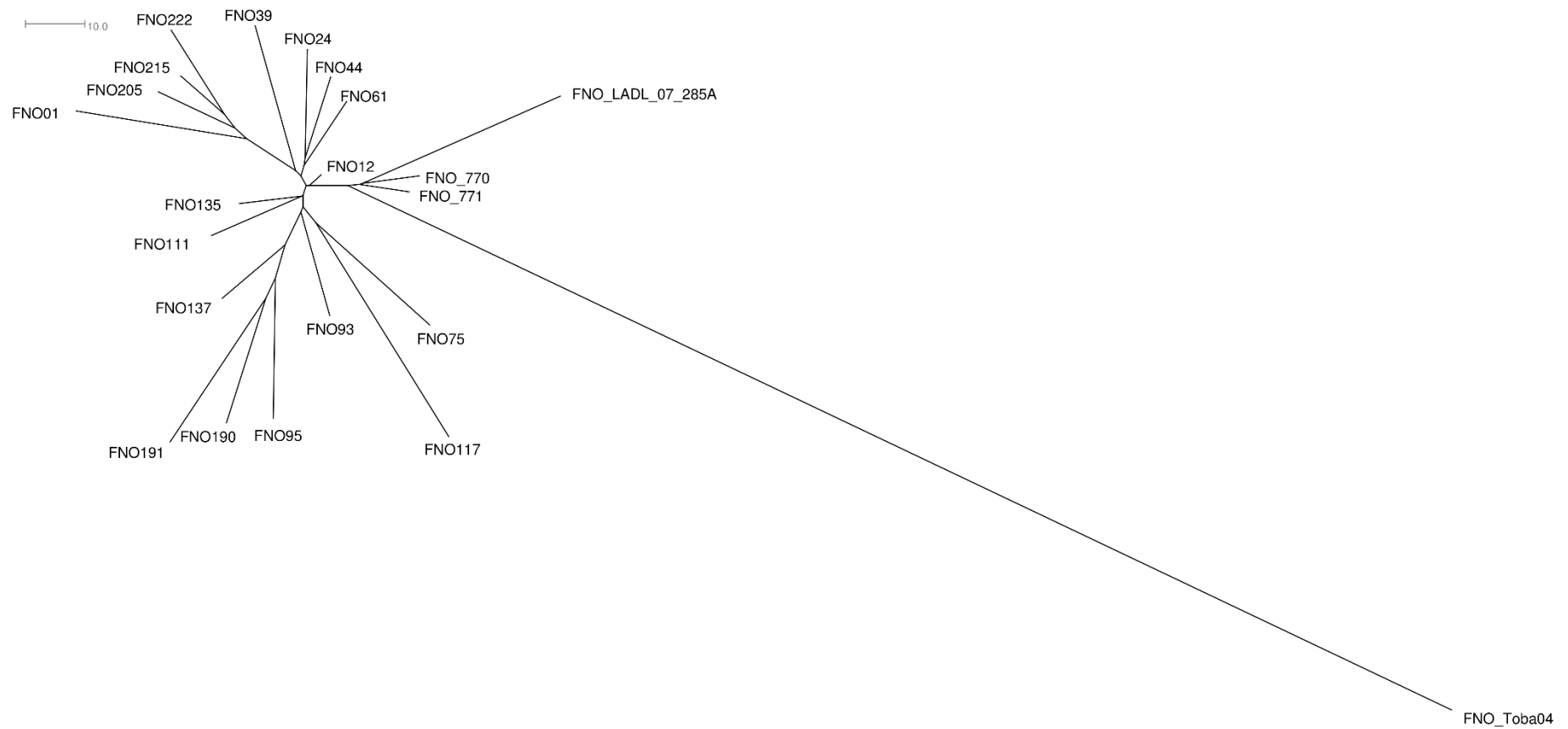


Figura 16: Relações filogenéticas das linhagens de *Francisella noantunensis* subsp. *orientalis*. Resultado obtido com base na abordagem gene a gene do genoma completo. O resultado indica a variabilidade das linhagens de *Francisella noantunensis* subsp. *orientalis*.

6.3. Genômica funcional

Devido ao fato da FNO ser uma bactéria de vida intracelular facultativa, ela passa por diversos estresses criados pela célula hospedeira, em especial por células de defesa, tais como macrófagos. Desse modo, para que ela sobreviva apropriadamente, culminando no processo patogênico, a mesma deve resistir a essas condições, que envolvem, estresses oxidativo e ácido, por exemplo. Portanto, buscamos caracterizar genes codificantes de proteínas que poderiam estar associadas com os processos de resistência a tais mecanismos de defesa do hospedeiro.

6.3.1. Estresse oxidativo

Com a exceção das bactérias anaeróbicas, o oxigênio é indispensável para todos os organismos, pois eles dependem da sua presença no ambiente para a produção de energia, na forma de ATP, durante um processo chamado de fosforilação oxidativa (Lushechak, 2001).

A fosforilação oxidativa, sendo num organismo eucarioto ou procaríoto (aeróbio estrito ou facultativo), está associada com a redução de moléculas de O_2 em H_2O (Lushechak, 2001). Contudo, apesar desse processo, normalmente, resultar na redução completa do O_2 , a sua redução univalente pode ocorrer, permitindo a geração de espécies reativas de oxigênio (ROS – do inglês, Reactive Oxygen Species) como o superóxido (O_2^-)¹, o peróxido de hidrogênio (H_2O_2) e o radical hidroxila (HO^*) (Hassett & Cohen, 1989).

Nessa situação, quando a geração de ROS se dá no contexto do metabolismo do próprio organismo, ela é referida como estresse oxidativo “endógeno”. Todavia, determinados organismos, como as bactérias, podem ser expostos a condições ambientais nas quais ROS exógenas são formadas, sendo, o exemplo mais notável desse estresse “exógeno”, o ataque de células fagocíticas à bactérias patogênicas (Hassett & Cohen, 1989).

A cadeia transportadora de elétrons das células fagocíticas, como macrófagos, é significativamente diferente daquela de outros tipos celulares uma vez que, nas células fagocíticas, cada um dos componentes que fazem parte dessa cadeia necessitam ser, individualmente, agregados à membrana, requerendo vários passos para a sua ativação. Ademais, esses fagócitos têm uma competência considerável de produzir (O_2^-)¹ por meio da transferência de um elétron proveniente de um NADPH para uma molécula de O_2 , sendo que, subsequente à geração de (O_2^-)¹, uma variedade de ROS

intermediárias são produzidas. Estas, por sua vez, são as que possuem o maior efeito antimicrobiano (Hassett & Cohen, 1989).

DNA, RNA, proteínas e os lipídeos são os alvos biológicos das ROS (Figura 17) e com relação aos danos que podem ser produzidos ao DNA, é possível que essas espécies afetem tanto as bases nitrogenadas quanto as moléculas de açúcar (desoxirribose). Isso pode originar quebras no arcabouço açúcar-fosfato, na formação de adutos de bases e açúcar, em ligações cruzadas do DNA com outras moléculas e em outras lesões que impeçam a replicação do DNA. Com relação às proteínas os danos que podem ser causados compreendem a redução de dissulfetos, oxidação de grupamentos sulfidríla, a reação de proteínas com aldeídos, a adução de resíduos de aminoácidos, a modificação de grupos prostéticos ou de *clusters* de metais, a ligação cruzada entre proteínas e a fragmentação de peptídeos. Adicionalmente, as ROS podem atacar os ácidos graxos poliinsaturados presentes na membrana citoplasmática, levando à peroxidação desses lipídeos e esse evento, por sua vez, reduz a fluidez da membrana, o que pode, ainda, resultar na liberação das proteínas integradas a ela (Cabiscol *et al.*, 2000).

As alterações supracitadas são consideradas deletérias para as células bacterianas e, por esse motivo, a sua sobrevivência depende da evolução de uma série de mecanismos de defesa (Cabiscol *et al.*, 2000), como: prevenção da geração das ROS, reparo de danos e impedimento da propagação das ROS, (Lushechak, 2001). Além disso, a resposta bacteriana ao estresse oxidativo não é estática, podendo ser regulada coordenadamente (Hassett & Cohen, 1989) por vários reguladores da expressão gênica.

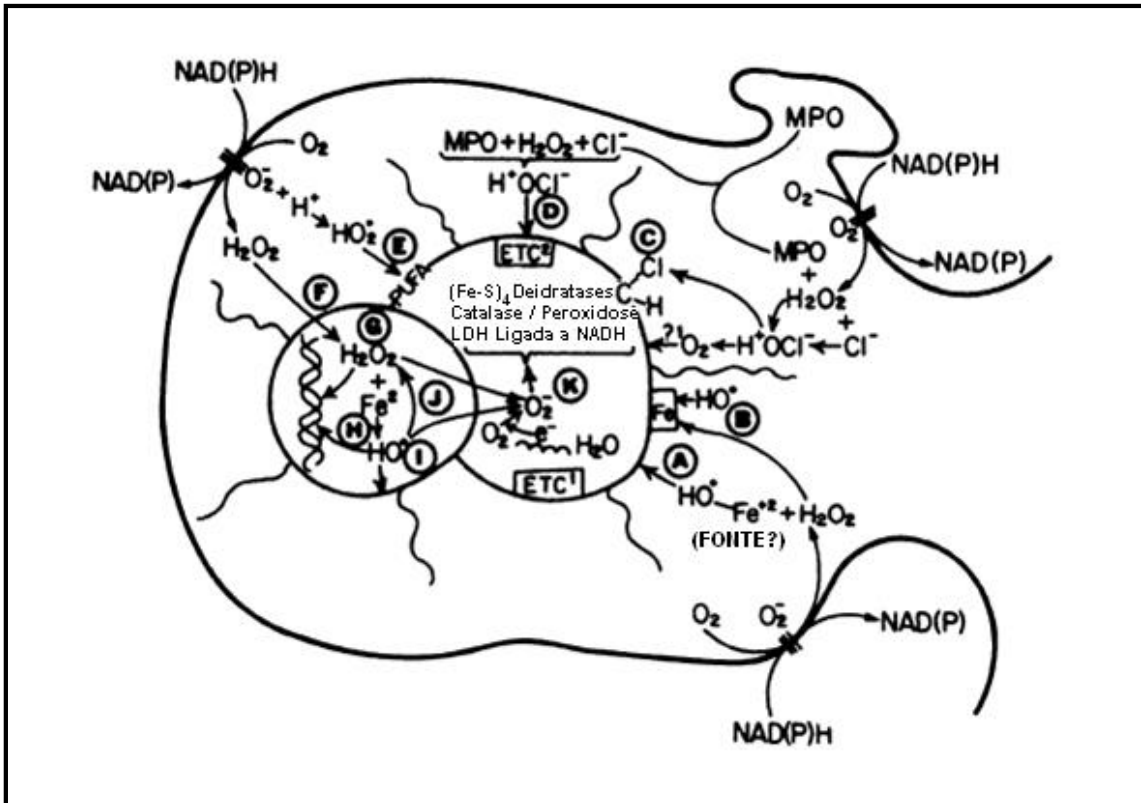


Figura 17: Esquema exemplificativo de como as espécies reativas de oxigênio (ROS) produzidas pelas células fagocíticas podem causar danos às bactérias patogênicas. (A) HO^\bullet extracelular pode danificar a membrana externa da bactéria; a fonte de Fe^{+2} , neste caso, ainda é desconhecida. (B) H_2O_2 extracelular pode reagir com proteínas, as quais estão integradas à superfície celular bacteriana, que apresentam Fe em sua estrutura para catalisar a síntese de HO^\bullet . (C) A formação de HOCl^- pode causar reações de cloração, levando a bactéria a ter problemas com o seu sistema de transporte celular. Além disso, HOCl^- pode reagir com H_2O_2 para formar $^1\text{O}_2$ (o que ocorre no caso do fagócito ser um eosinófilo), o que pode danificar a membrana citoplasmática do microrganismo. (D) HOCl^- pode causar a oxidação de grupos prostéticos Fe-S, citocromos e enzimas importantes que atuam na cadeia transportadora de elétrons (ETC – do inglês, *electron transport chain*). (E) HO_2^\bullet pode causar a oxidação de ácidos graxos poliinsaturados (PUFA – do inglês, *polyunsaturated fatty acids*) da membrana bacteriana. (F) H_2O_2 pode interferir no potencial ΔP da membrana citoplasmática da bactéria, levando-a a ter problemas com o seu sistema de transporte celular. (G) H_2O_2 pode causar a quebra do DNA, levando à mutagenese e à morte celular. (H) H_2O_2 pode reagir com o Fe celular, formando HO^\bullet , o qual pode danificar criticamente o DNA; H_2O_2 pode oxidar a parede celular e componentes da membrana externa bacterianas. (I) HO^\bullet pode danificar extensamente a membrana interna da bactéria. (J) HO^\bullet pode reagir com o H_2O_2 em excesso para formar O_2^- . (K) O_2^- gerado pelo metabolismo bacteriano pode inativar enzimas importantes como a catalase/peroxidase, a lactato desidrogenase ligada a NADH (LDH – do inglês, *lactate dehydrogenase*) e a α,β -diidroxiácido desidratase. **ADAPTADO DE:** Hassett & Cohen, 1989.

Um exemplo de proteínas reguladoras que participam desse tipo de resposta são os reguladores transcricionais tipo LysR (LTTR – do inglês LysR-type transcriptional regulator) (Sridhar *et al.*, 2012), os quais formam um grupo bem caracterizado, altamente conservado e onipresente entre bactérias. LTTRs são encontrados nas diferentes subdivisões de proteobactérias, sendo que nas linhagens brasileiras de

Francisella noatunensis subsp. *orientalis* – uma γ -proteobactéria – foram encontrados 7 genes que codificam proteínas pertencentes a essa família (Tabela 9).

Tabela 9: LTTRs encontradas na linhagem FNO12

Produto	Localização no genoma de FNO12 - (locus tag)
OxyR - Oxidative stress transcriptional regulator	FNO12_1067
Regulador transcricional da família LysR	FNO12_0026
Regulador transcricional da família LysR	FNO12_0071
Regulador transcricional da família LysR	FNO12_0078
Regulador transcricional da família LysR	FNO12_0581
Regulador transcricional da família LysR	FNO12_1207
Regulador transcricional da família LysR	FNO12_1739

Desses, um tem como produto a proteína OxyR, a qual controla a regulação de genes que medeiam a resposta ao estresse oxidativo (H_2O_2) (Schell, 1993), ativando a transcrição de genes alvo após a oxidação intramolecular de cisteínas situadas em locais estratégicos (Figura 18) (Vázquez-Torres, 2012). Os genes *katG* e *ahpC*, também encontrados nas linhagens de FNO, fazem parte do regulon de OxyR. O primeiro codifica a enzima catalase-peroxidase KatG que possui atividade catalática e peroxidática ao sequestrar moléculas de H_2O_2 , protegendo assim as células da sua toxicidade (Jittawuttipoka *et al.*, 2009). Já o gene *ahpC* codifica a subunidade C da enzima alquil hidroperóxido redutase, que reduz os peróxidos orgânicos (Chen *et al.*, 1998). Um fato interessante é que para vários microrganismos, estes dois genes têm padrões de expressão compensatórias uma vez que a inativação de catalase leva a um aumento no nível de AhpC e um aumento nos níveis de expressão da catalase parecem reduzir os níveis de AhpC (Charoenlap *et al.*, 2005). Além disso, a enzima alquil hidroperóxido redutase é capaz de sequestrar, de maneira mais eficiente, H_2O_2 em baixas concentrações. Desse modo, Ahp parece ser o responsável pela eliminação primária de H_2O_2 , enquanto a catalase se torna essencial na presença de altas concentrações de H_2O_2 (Panek & O'Brian, 2004).

Outras proteínas LTTRs são importantes para a resposta ao estresse oxidativo. Assim, como em FNOToba04 e FNOLADL-07-285A, foram encontrados seis genes codificadores desse tipo de proteínas (Tabela 9) ao passo que existem oito proteínas LysR em *F. tularensis* subsp. *tularensis* SCHU S4, cinco em *F. tularensis* subsp. *holarctica* OSU18 e nenhuma em linhagens de *F. tularensis* subsp. *mediasiatica*, a qual apresenta uma virulência moderada em relação a outras linhagens de *F. tularensis* (Sridhar *et al.*, 2012).

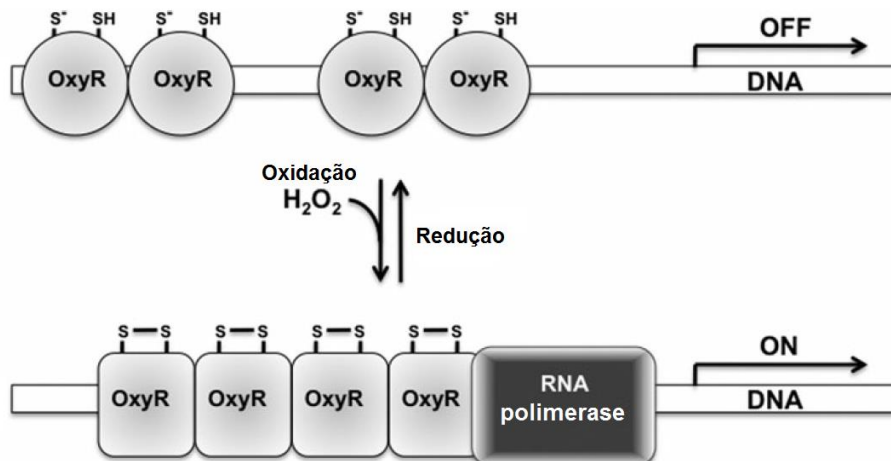


Figura 18: Detecção de peróxido de hidrogênio (H_2O_2) por OxyR. OxyR reduzido se liga ao promotor de genes alvo, impedindo a transcrição dos mesmos (acima). A oxidação do grupo tiolato ($-S^-$) da Cys¹⁹⁹ de OxyR por H_2O_2 produz um derivado do ácido sulfênico ($-SOH$) que ataca a cadeia lateral tiol de Cys²⁰⁸. A ligação intramolecular disulfeto resultante leva à alterações estruturais de OxyR, permitindo o recrutamento da RNA polimerase e a ativação da transcrição (embaixo). **ADAPTADO DE:** Vázquez-Torres, 2012

Portanto, a análise dos genes relacionados ao estresse oxidativo aqui descritos não evidencia diferenças entre as linhagens de *Francisella noatunensis* subsp. *orientalis*, uma vez que não foram encontradas diferenças na quantidade ou existência desses genes. Sob a ótica da genômica isso leva a crer que o comportamento das linhagens sob condições desse estresse deva ser o mesmo.

6.3.2. Estresse térmico

Temperaturas apenas moderadamente acima e abaixo do ponto de crescimento ótimo representam um desafio para a sobrevivência de qualquer ser vivo, tornando o calor e o frio uns dos principais agentes causadores de estresse (Phadtare & Severinov, 2010; Richter et al., 2010). A resposta ao choque térmico, portanto, é ativada por um aumento de poucos graus Celsius ($^{\circ}C$) na temperatura, o que pode causar, dentre outras coisas, desdobramento, emaranhamento e agregação inespecífica de proteínas. Esses efeitos, juntos, são capazes de impedir o progresso do ciclo celular, resultando, por sua vez, na estagnação do crescimento do organismo. Ademais, dependendo da duração e da severidade do estresse térmico a que um organismo é submetido, é possível que o acúmulo de tais efeitos deletérios resulte na morte celular (Richter et al., 2010).

Por outro lado, a resposta ao choque frio consiste em uma série de modificações adaptativas que vão desde alterações na composição da membrana plasmática a

alterações no perfil de proteínas da célula com o intuito de evitar: (i) a redução da fluidez da membrana, o que afeta funções celulares como o transporte ativo e a secreção de proteínas; (ii) a redução da função do ribossomo, o que afeta a tradução, principalmente, de proteínas não relacionadas à resposta ao choque frio; (iii) o enovelamento ineficiente de proteínas; e (iv) modificações na estrutura dos ácidos nucleicos, as quais podem ser a estabilização das estruturas secundárias do RNA e do DNA, que afeta os processos de transcrição e tradução; e o aumento da helicoidização negativa do DNA, que afeta os processos de replicação, transcrição e recombinação (Phadtare & Severinov, 2010).

Todos os organismos respondem ao aumento e à redução repentina de temperatura com o aumento da expressão de um grupo de proteínas altamente conservadas denominadas proteínas de choque térmico – HSPs, do inglês *heat shock proteins* – e proteínas de choque frio – CSPs, do inglês *cold shock proteins*, respectivamente. Essa resposta, por sua vez, é controlada por uma combinação de fatores sigma alternativos, os quais reconhecem promotores de genes de choque térmico e choque frio específicos, e reguladores transcricionais, que agem como repressores ou ativadores de certos genes de choque térmico e choque frio ao se ligarem à região promotora de tais genes específicos (Ehira *et al.*, 2009).

Assim como para *F. tularensis* LVS, a análise *in silico* do genoma de FNO12 permitiu identificar todos os genes envolvidos na síntese do cerne da holoenzima RNA polimerase (RNAP): *rpoA1* e *rpoA2* que codificam as subunidades α ; *rpoB* e *rpoC* que codificam as subunidades β e β' ; e *rpoZ* que codifica a subunidade ω . Além disso, foram identificados os genes *rpoD* que codifica o fator σ^{70} e *rpoH* que codifica o fator sigma alternativo σ^{32} . Desse modo, FNO pertence a um grupo relativamente pequeno de espécies bacterianas que expressam apenas um único fator sigma alternativo. O controle da transcrição por *rpoH* parece ser muito específico para as proteínas de choque térmico uma vez que ela parece ativar a transcrição dos genes *groL*, *groS*, *dnaK* e *dnaJ* também encontrados nas FNOs (Grall *et al.*, 2009).

Também foram identificados outros genes que codificam proteínas relacionadas à resposta ao choque térmico, como *grpE* que faz parte do operon *dnaJ-dnaK-grpE*, o gene codificador da proteína HSP15 e mais dois genes que foram anotados somente o produto como sendo HSP (Tabela 10). Essas proteínas garantem um correto enovelamento das proteínas e são essenciais para resistir ao estresse através da redução do dano causado pelo acúmulo de proteínas degradadas e enoveladas de maneira errada (Grall *et al.*, 2009).

Adicionalmente, elas possuem dois genes codificadores de proteínas de choque frio, dentre eles o *cspA* cujo gene homólogo codifica a principal e mais abundante

proteína de choque frio em *E. coli* (Phadtare & Severinov, 2010). No entanto, mesmo estando presentes em todas as FNOs, os genes codificadores de HSPs e CSPs tiveram, possivelmente, um papel importante na adaptação dessa espécie ao seu hospedeiro uma vez que a variação na temperatura da água pode ser considerada uma condição adversa. Ademais, a presença de proteínas de choque térmico são importantes para a sobrevivência da bactéria no ambiente intrafagossomal, sendo, assim, essencial para a patogenicidade desse microrganismo.

Tabela 10: Genes de choque térmico e frio encontrados em FNO12

	Gene ou locus tag	Produto
Proteínas de choque térmico	FNO12_0246	<i>Heat shock protein</i>
	FNO12_0385	<i>Heat shock protein</i>
	FNO12_0808	<i>Heat shock protein 15 (HSP15)</i>
	<i>groL</i>	<i>60 kDa chaperonin</i>
	<i>groS</i>	<i>10 kDa chaperonin</i>
	<i>dnaJ</i>	<i>Chaperone protein</i>
	<i>dnaK</i>	<i>Chaperone protein</i>
Proteínas de choque frio	<i>grpE</i>	<i>grpE / Heat Shock protein</i>
	<i>cspA</i>	<i>Cold shock protein CspA</i>
	FNO12_0817	<i>Cold shock protein</i>

6.3.3. Estresse ácido

Um dos fatores que contribuem para a sobrevivência da bactéria em ambientes cujo pH é baixo é a resistência ao ácido e um exemplo de ambiente que oferece esse tipo de adversidade às bactérias é o fagolisossomo no interior do macrófago (Figura 19). Esse tipo de resistência, por sua vez, contribui para a virulência de um organismo patogênico (Cotter & Hill, 2003) como *Francisella*. Isso é verdade uma vez que os mecanismos que conferem esse fenótipo ao microrganismo possibilitam a geração de uma resposta aos impactos negativos da redução do pH citoplasmático, os quais podem incluir danos estruturais à membrana celular e à macromoléculas essenciais como o DNA, algumas proteínas e a perda da atividade de enzimas relativamente sensíveis ao ácido que atuam na via glicolítica, afetando severamente a habilidade da célula de produzir ATP (Cotter & Hill, 2003).

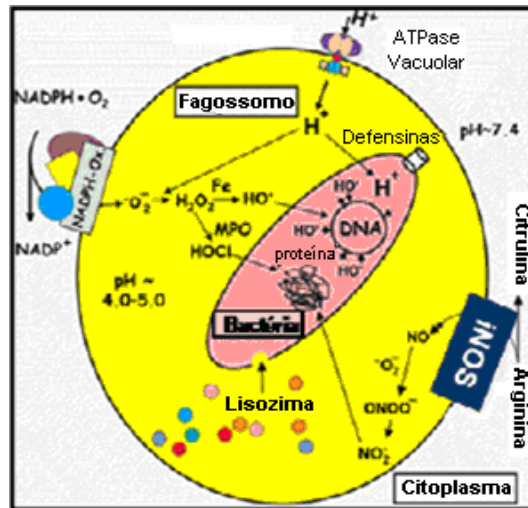


Figura 19: Diferentes condições de estresse encontradas por bactérias patogênicas no ambiente intrafagossômico. **ADAPTADO DE:** Pacheco, 2010.

Em geral, a resistência ao ácido é dada por três principais estratégias: (I) mudanças na composição da membrana citoplasmática do microrganismo, (II) manutenção do pH interno da célula e (III) prevenção e/ou reparo de danos causados a componentes celulares essenciais (Choi *et al.*, 2000). Ela é devida à combinação de estratégias constitutivas e induzíveis que resultam em ações como a remoção de prótons (H⁺) do ambiente intracelular, alcalinização do ambiente externo à célula, modificações na composição do envelope celular, produção de proteínas de choques gerais e chaperonas, expressão de reguladores transcricionais e respostas às mudanças na densidade da célula (Cotter & Hill, 2003).

Em FNO foram encontrados os genes *dnaK*, *groL*, *clpP*, e *clpX* (Clp ATPases) (Tabela 11) que codificam chaperoninas e proteases que atuam na proteção de proteínas ou na sua degradação caso estas estejam danificadas. Com a função de atuar no reparo do DNA que, neste caso pode ser feito por meio da excisão de bases lesadas / mal pareadas ou no recomeço de forquilha de replicação, foram encontrados os genes *recA*, *uvrA*, *uvrB* e *uvrC* (Tabela 11). Já com a função de alterar o envelope celular, podendo proteger as células por meio da mudança da sua arquitetura, composição e estabilidade, foi encontrado o gene *ffh* (Tabela 11). Por fim, também foram encontrados os genes *ureAB* e *ureD* (Tabela 11), os quais fazem parte do sistema urease que atua no aumento do pH intracelular por meio da produção de álcali (Cotter & Hill, 2003).

Tabela 11: Genes relacionado ao estresse ácido encontrados em FNO12

Gene ou locus	Produto
tag	
<i>dnaK</i>	<i>Chaperone protein</i>
<i>clpP</i>	<i>ATP-dependent Clp protease proteolytic subunit</i>
<i>clpX</i>	<i>ATP-dependent Clp protease ATP-binding subunit ClpX</i>
<i>recA</i>	<i>Protein RecA</i>
<i>uvrA</i>	<i>UvrABC system protein A</i>
<i>uvrB</i>	<i>UvrABC system protein B</i>
<i>uvrC</i>	<i>UvrABC system protein C</i>
<i>ffh</i>	<i>Signal recognition particle, subunit Ffh</i>
FNO12_0196	<i>Urease alpha subunit</i>
<i>ureAB</i>	<i>Urease beta and gamma subunits</i>
<i>ureD</i>	<i>Urease accessory protein</i>
<i>groL</i>	<i>60 kDa chaperonin</i>

Logo, pode-se dizer que a identificação dos genes relacionados ao estresse ácido, bem como os relacionados a outros tipos de estresse, constitui uma etapa importante e obrigatória no conhecimento do patógeno e atende a um fim que é o de buscar alvos para trabalhos futuros que podem culminar, por exemplo, na obtenção de uma linhagem vacinal.

7. CONCLUSÕES

A caracterização dos genomas das linhagens de *Francisella noatunensis* subsp. *orientalis* isoladas de surtos ocorridos no Brasil mostrou que os genomas são muito semelhantes em nível de nucleotídeos, conteúdo G+C, número de CDSs, número de pseudogenes, RNAs não codificantes. Esses resultados são, também, muito próximos dos disponíveis para os genomas FNOLADL-07-287A e FNOToba04. Avaliações feitas com Mauve e Gegenees também atestaram esses resultados e mostraram que as linhagens são sintênicas e muito próximas em análise de filogenia baseada em similaridade de nucleotídeos.

Ademais, análises de polimorfismo realizadas no BIGSdb indicaram que as linhagens de FNO, especialmente as brasileiras, são clonais e que provavelmente a chegada da doença ao país se deu pela entrada de uma única linhagem e desde então ela tem se diversificado enquanto é disseminada nas fazendas.

Avaliações acerca da plasticidade genômica predisseram oito ilhas genômicas que possivelmente contêm genes responsáveis pelo seu fenótipo. Alguns desses, inclusive, estão relacionados à sobrevivência em mais de um tipo de estresse. Isso qualifica tais genes como candidatos a um estudo mais aprofundado para o desenvolvimento de drogas que podem ter um retorno positivo no controle de surtos e redução de perdas econômicas no cultivo de tilápias.

Desse modo, o presente trabalho acrescenta informações ao conhecimento disponível sobre o patógeno e serve como base para que o Ministério da Pesca possa determinar condutas de vigilância sobre a Franciselose. Além de criar possibilidades de estudos complementares sobre o patógeno.

8. PERSPECTIVAS

- Os dados gerados por este trabalho abrem perspectivas para estudos em novas ômicas;
- O sequenciamento dos genomas de FNO possibilitam que mais estudos, como análises do pangenoma, sejam realizados;
- O presente trabalho pode servir como base para a escolha de genes candidatos (relacionados à virulência) para a construção de mutantes e avaliar a persistência da infecção e da virulência *in vitro* e *in vivo*;
- Caso apresente um fenótipo atenuado, testar a linhagem mutante como uma possível vacina contra a franciselose;

9. REFERÊNCIAS BIBLIOGRÁFICAS

- Ahlinder J, Öhrman C, Svensson K, et al. Increased knowledge of Francisella genus diversity highlights the benefits of optimised DNA-based assays. *BMC Microbiology* 2012;12:220.
- Barns, SM., Grow, C.C., Okinaka, R.T., Keim, P. and Kuske, C.R. (2005) Detection of diverse new Francisella-like bacteria in environmental samples. *Appl Environ Microbiol* 71, 5494–5500.
- Baron, GS. and Nano, F. E. MglA and MglB are required for the intramacrophage growth of Francisella novicida. *Molecular Microbiology*, (1998) 29: 247–259.
- Birkbeck TH, Feist SW, Verner-Jeffreys DW. Francisella infections in fish and shellfish. *J Fish Dis* 2011; 34(3):173-87.
- Bozano, GLN; Cyrino, JEP. Produção intensiva de peixes em tanques-rede e gaiolas: estudo de casos. *Panorama da Aquicultura*, v. 56, n. 6, p. 36-41, 1999.
- Brudal E, Ulanova LS, O. Lampe E, Rishovd A-L, Griffiths G, Winther-Larsen HC. Establishment of Three Francisella Infections in Zebrafish Embryos at Different Temperatures. Morrison RP, ed. *Infection and Immunity* 2014;82(6):2180-2194.
- Byers DM, Gong H. Acyl carrier protein: structure-function relationships in a conserved multifunctional protein family. *Biochem Cell Biol.* 2007. 85 (6) : 649-62.
- Cabedo H, Macián F, Villarroya M, Escudero JC, Martínez-Vicente M, Knecht E, Armengod ME. The Escherichia coli trmE (mnmE) gene, involved in tRNA modification, codes for an evolutionarily conserved GTPase with unusual biochemical properties. *The EMBO Journal.* 1999, 18 (24) : 7063-7076.
- Cabiscol, E; Tamarit, J; Ros, J. Oxidative stress in bacteria and protein damage by reactive oxygen species. 2000; *Internati. Microbiol.*, 3: 3-8.
- Callou, ABF. Povos do mar: herança sociocultural e perspectivas no Brasil. *Cienc. Cult.* 2010, vol.62, n.3, pp. 45-48. ISSN 2317-6660.
- Carriço JA, Sabat AJ, Friedrich AW, Ramirez M; ESCMID Study Group for Epidemiological Markers (ESGEM). *Bioinformatics in bacterial molecular epidemiology and public health: databases, tools and the next-generation sequencing revolution.* *Euro Surveill.* 2013;18(4):20382.
- Charoenlap N, Eiamphungporn W, Chauvatcharin N, Utamapongchai S, Vattanaviboon P, Mongkolsuk S. OxyR mediated compensatory expression between ahpC and katA and the significance of ahpC in protection from hydrogen peroxide in Xanthomonas campestris. *FEMS Microbiol Lett.* 2005; 1;249(1):73-8.

- Chen L, Xie QW, Nathan C. Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol Cell*. 1998. 1 (6) : 795-805.
- Choi, SH; Baumler, DJ.; Kaspar, CW. Contribution of dps to acid stress tolerance and oxidative stress tolerance in *Escherichia coli* O157:H7. *Appl. Environ. Microbiol*. 2000, 66 (9): 3911-3916.
- Colguhoun DJ, Duodu S. *Francisella* infections in farmed and wild aquatic organisms. *Vet Res* 2011; 42:47.
- Cotter, P.D.; Hill, C. Surviving the acid test: responses of Gram-positive bacteria to low pH. *Microbiol. Mol. Biol. Rev.* 2003. 67 (3): 429-453.
- Croucher NJ, Harris SR, Grad YH, Hanage WP. Bacterial genomes in epidemiology—present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2013; 368 (1614) : 20120202.
- Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* (2010) 5: e11147.
- Ehira, S; Teramoto, H; Inui, M; Yukawa, H. Regulation of *Corynebacterium glutamicum* heat shock response by the extracytoplasmic-function sigma factor SigH and transcriptional regulators HspR and HrcA. *J. Bacteriol*. 2009; 191 (9): 2964-2972.
- El-Metwally S, Hamza T, Zakaria M, Helmy M. Next-Generation Sequence Assembly: Four Stages of Data Processing and Computational Challenges. 2013 *PLoS Comput Biol* 9(12): e1003345.
- Eschenburg S, Kabsch W, Healy ML, Schonbrunn E. A new view of the mechanisms of UDP-N-acetylglucosamine enolpyruvyl transferase (MurA) and 5-enolpyruvylshikimate-3-phosphate synthase (AroA) derived from X-ray structures of their tetrahedral reaction intermediate states. *J Biol Chem*. 2003 5 ; 278 (49) : 49215-22.
- Grad, Y. H.; Lipsitch, M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Gen. Biol*. 2014, 15:538
- Grall N, Livny J, Waldor M, Barel M, Charbit A, Meibom KL. Pivotal role of the *Francisella tularensis* heat-shock sigma factor RpoH. *Microbiology*. 2009; 155(Pt 8):2560-2572.
- Gonçalves, A. A.; Passos, M. G.; Biedrzycki, A. Tendência do consumo de pescado na cidade de Porto Alegre: um estudo através de análise de correspondência. *Estudos Tecnológicos*, v. 4, n. 1, p. 21-36, 2008.

- Gonçalves, A. A.; Passos, M. G.; Biedrzucki, A. Percepção do consumidor com relação à embalagem de pescado: estudo de caso com os alunos do curso de Engenharia de Alimentos. *Estudos Tecnológicos*, v. 5, n. 1, p. 14-32, 2009.
- Harismendy O, Ng PC, Strausberg RL, et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*. 2009; 10 (3) : R32.
- Hare RF, Hueffer K. Francisella novicida ilha de patogenicidade codificado proteínas foram secretado durante a infecção de células-macrófagos Like. Kwaik YA, ed. *PLoS ONE* . 2014; 9 (8): e105773.
- Hassett, DJ; Cohen, MS. Bacterial adaptation to oxidative stress: implications for pathogenesis and interaction with phagocytic cells. *The FASEB Journal*. 1989; 3: 2574-2582.
- Jones, B.D.; Faron, M.; Rasmussen, J. A.; Fletcher, J.R.; Uncovering the components of the Francisella tularensis virulence stealth strategy. *Front. Cell. Infect. Microbiol*. 2014.
- Jittawuttipoka, T.; Buranajitpakorn, S.; Vattanaviboon, P.; Mongkolsuk.; S. The Catalase-Peroxidase KatG Is Required for Virulence of Xanthomonas campestris pv. campestris in a Host Plant by Providing Protection against Low Levels of H₂O₂. *Journal Of Bacteriology*. 2009, 191 (23): 7372–7377
- IBGE (Instituto Brasileiro de Geografia e Estatística). *Prod. Pec. munic.*, Rio de Janeiro 2013; v. 41, p.1-108.
- Itaya, M. Isolation and characterization of a second RNase H (RNase HII) of Escherichia coli K-12 encoded by the rnhB gene. *Proc. Natl. Acad. Sci. U.S.A.* 1990, 87(21): 8587–8591.
- Kato K. Impact of the next generation DNA sequencers. *International Journal of Clinical and Experimental Medicine* 2009;2(2):193-202.
- Kaur R, Malik C.P. (2013) Next Generation Sequencing: A revolution in gene sequencing. *CIBTech Journal of Biotechnology* 2 (4): 1-20.
- Kingry LC, Petersen JM. Comparative review of Francisella tularensis and Francisella novicida. *Frontiers in Cellular and Infection Microbiology* 2014; 4:35.
- Kubitza, F. Antecipando-se às doenças na tilapicultura. *Panorama da Aquicultura*, v. 15, n. 89, p. 15-23, 2005.
- Kugeler, K.J., Mead, P.S., McGowan, K.L., Burnham, J.M., Hogarty, M.D., Ruchelli, E., Pollard, K., Husband, B. et al. (2008) Isolation and characterization of a novel Francisella sp. from human CSF and Blood. *J Clin Microbiol* 46, 2428–2431.

- Kuznetsova E1, Proudfoot M, Gonzalez CF, Brown G, Omelchenko MV, Borozan I, Carmel L, Wolf YI, Mori H, Savchenko AV, Arrowsmith CH, Koonin EV, Edwards AM, Yakunin AF. Genome-wide analysis of substrate specificities of the Escherichia coli haloacid dehalogenase-like phosphatase family. *J Biol Chem.* 2006 24; 281(47) : 36149-61.
- Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 2010; 11:595.
- Larsen MV, Cosentino S, Rasmussen S, et al. Multilocus Sequence Typing of Total-Genome-Sequenced Bacteria. *Journal of Clinical Microbiology* 2012;50(4):1355-1361.
- Larsson P, Elfsmark D, Svensson K, et al. Molecular Evolutionary Consequences of Niche Restriction in *Francisella tularensis*, a Facultative Intracellular Pathogen. Guttman DS, ed. *PLoS Pathogens* 2009;5(6):e1000472.
- Li Y, Du J, Zhang P, Ding J. Crystal structure of human copper homeostasis protein CutC reveals a potential copper-binding site. *J Struct Biol.* 2010; 169 (3) : 399-405.
- Lommatzsch J, Templin MF, Kraft AR, Vollmer W, Holtje JV. Outer Membrane Localization of Murein Hydrolases: MltA, a Third Lipoprotein Lytic Transglycosylase in *Escherichia coli*. *Jornal of Bacteriology.* 1997; V: 179, n. 17, p. 5465–5470.
- Lushechak, V.I. Oxidative stress and mechanisms of protection against it in bacteria. *Biochemistry.* 2001; 66 (5) : 476-489.
- Ma L-S, Narberhaus F, Lai E-M. IcmF Family Protein TssM Exhibits ATPase Activity and Energizes Type VI Secretion. *The Journal of Biological Chemistry.* 2012; 287(19) : 15610-15621.
- Maiden MCJ, Jansen van Rensburg MJ, Bray JE, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature reviews. Microbiology* 2013.
- Metzker, M. L. Emerging technologies in DNA sequencing. *Genome Res.* 2005; 15, 1767–1776.
- Mizote T, Nakayama H. The thiM locus and its relation to phosphorylation of hydroxyethylthiazole in *Escherichia coli*. *J Bacteriol.* 1989. 171 (6) : 3228–3232.
- Mian, G. F.; Godoy, D. T.; Leal, C. A.G. et al. Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. *Vet. Microbiol.* v. 136, n. 1-2, p. 180-183, 2008.
- Marquardt JL1, Siegele DA, Kolter R, Walsh CT. Cloning and sequencing of *Escherichia coli* murZ and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase. *J Bacteriol.* 1992; 174 (17) : 5748-52.

- Niebylski, M.L., Peacock, M.G., Fisher, E.R., Porcella, S.F. and Schwan, T.G. (1997) Characterization of an endosymbiont infecting wood ticks, *Dermacentor andersoni*, as a member of the Genus *Francisella*. *Appl Environ Microbiol* 63, 3933–3940.
- Nishino H, Biogenesis of cocarboxylase in *Escherichia coli*. Partial purification and some properties of thiamine monophosphate kinase. *J Biochem.* 1972. 72 (5) : 1093-100.
- Ottem KF, Nylund A, Karlsbakk E, Fris-Moller A, Kamaishi T. Elevation of *Francisella philomiragia* subsp. *noatunensis* Mikalsen et al. (2007) to *Francisella noatunensis* comb. nov. [syn. *Francisella piscicida* Ottem et al. (2008) syn. nov.] and characterization of *Francisella noatunensis* subsp. *orientalis* subsp. nov., two important fish pathogens. *Journal of Applied Microbiology* (2009); 106:1231–1243.
- Oyston PCF, Sjöstedt A, Titball RW. Tularaemia: bioterrorism defence renews interest in *Francisella tularensis*. *Nature Reviews Microbiology* (2004), 967-978.
- Pacheco, LGC. Caracterização do exoproteoma basal e variante de *Corynebacterium pseudotuberculosis* e a relação com a virulência reduzida de uma linhagem mutante para um fator sigma alternativo. 2010. (Tese doutorado em bioquímica e imunologia) – Instituto de Ciências Biológicas – Universidade Federal de Minas Gerais. Belo Horizonte, 2010.
- Panek, HR & O’Brian, ME. KatG Is the Primary Detoxifier of Hydrogen Peroxide Produced by Aerobic Metabolism in *Bradyrhizobium japonicum*. 2004; 184 (23) 7874–7880.
- Parkhill J, Wren BW, Bacterial epidemiology and biology - lessons from genome sequencing. *Gen. Biol.* 2011, 12 : 230.
- Phadtare S, Severinov K. RNA remodeling and gene regulation by cold shock proteins. *RNA Biology.* 2010;7(6):788-795.
- Petersen LA, Downs DM. Identification and characterization of an operon in *Salmonella typhimurium* involved in thiamine biosynthesis. *J Bacteriol.* 1997. 179 (15) : 4894-900.
- Pereira, U. P.; Mian, G. F.; Oliveira, I. C. M. et al. Genotyping of *Streptococcus agalactiae* strains isolated from fish, human and cattle and their virulence potential in Nile tilapia. *Vet. Microbiol.*, v. 140, n. 1-2, p.186-192, 2010.
- Reddick JJ, Nicewonger R, Begley TP. Mechanistic studies on thiamin phosphate synthase: evidence for a dissociative mechanism. *Biochemistry.* 2001. 28; 40 (34) : 10095-102.
- Richter, K.; Haslbeck, M.; Buchner, J. The heat shock response: life on the verge of death. *Molecular Cell.* 2010; 22; 40 (2) : 253-66.

- Rodrigues, E. Pesquisa de Aeromonas spp. em tilápia (*Oreochromis niloticus*), cultivada no Estado do Rio de Janeiro – Brasil: Isolamento, identificação de espécies e avaliação da sensibilidade antimicrobiana. 2007. 210f. Tese (Doutorado) – Faculdade de Veterinária, Universidade Federal Fluminense, Niterói.
- Rohmer L, Fong C, Abmayr S, et al. Comparison of Francisella tularensis genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biology*. 2007; 8 (6) :R102.
- Santos, CAML. Qualidade do Pescado. In: Gonçalves, A. A. Tecnologia do Pescado: Ciência, Tecnologia, Inovação e Legislação. São Paulo: Editora Atheneu, 2011. p. 95-105.
- Schell, MA. Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol.* 1993; 47:597-626.
- Silva, JV; Cardoso, J. R.; Sousa, M. M. Qualidade sensorial de peixes comercializados em mercados públicos de Teresina-PI. *Rev. Hig. Alimentar*, v. 27, n. 3, p. 36-40, 2013.
- Sjödin A, Svensson K, Öhrman C, et al. Genome characterisation of the genus Francisella reveals insight into similar evolutionary paths in pathogens of mammals and fish. *BMC Genomics* 2012;13:268.
- Soto E, Hawke JP, Fernandez D, Morales JA. Francisella sp., an emerging pathogen of tilapia, oreochromis niloticus (L.), in Costa Rica. *J Fish Dis* 2009; 32 (8):713-22.
- Soto E, Fernandez D, Thune R, Hawke JP. Interaction of Francisella asiatica with Tilapia (*Oreochromis niloticus*) Innate Immunity. *Infection and Immunity* 2010;78(5):2070-2078.
- Soto E, Revan F. Culturability and persistence of Francisella noatunensis subsp. orientalis (syn. Franciesella asiática) in sea-and freshwater microcosms. *Micro Ecol* 2012; 63(2):398-404.
- Soto E, Illanes O, Hilchie D, Morales Ja, Sunyakumthorn P, Hawke JP, Goodwin AE, Riggs A, Yanong RP, Pouder DB, Francis-Floyd R, Arauz M, Bogdanovic L, Castillo-Alcala. Molecular and immunohistochemical diagnosis of Francisella noatunensis subsp. orientalis from formalin-fixed, paraffin-embedded tissues. *J Vet Diagn Invest* 2012a; 24(5):840-5.
- Soto E, Abrams SB, Revan F. Effects of temperature and salt concentration on Francisella noatunensis subsp. Orientalis infections in Nile tilapia Oreochromis niloticus. *Dis Aquat Organ* 2012b; 101(3):217-23.

- Soto E, Griffin M, Wiles J, Hawke J. P. Genetic analysis and antimicrobial susceptibility of *Francisella noatunensis* subsp. *orientalis* (syn. *F. asiatica*) isolates from fish. *Vet. Mic.* 2012c; 154:407-412.
- Soto E, Kidd S, Mendez S, Marancik D, Revan F, Hiltchie D, Camus A. *Francisella noatunensis* subsp. *orientalis* pathogenesis analyzed by experimental immersion challenge in Nile tilapia, *Oreochromis niloticus* (L.). *Vet Microbiol* 2013; 164(1-2):77-84.
- Sridhar S, Sharma A, Kongshaug H, Nilsen F, Jonassen I. Whole genome sequencing of the fish pathogen *Francisella noatunensis* subsp. *orientalis* Toba04 gives novel insights into *Francisella* evolution and pathogenicity. *BMC Genomics* 2012; 13:598.
- Tan J, Jakob U, Bardwell JC. Overexpression of two different GTPases rescues a null mutation in a heat-induced rRNA methyltransferase. *J Bacteriol.* 2002; 184 (10) : 2692-8.
- Tanous C, Soutourina O, Raynal B, Hullo MF, Mervelet P, Gilles AM, Noirot P, Danchin A, England P, Martin-Verstraete I. The CymR regulator in complex with the enzyme CysK controls cysteine metabolism in *Bacillus subtilis*. *The Journal of Biological Chemistry* 2008, 283 (51): 35551-35560.
- Vázquez-Torres, A. Redox active thiol sensors of oxidative and nitrosative stress. *Antioxidants & Redox Signaling.* 2012; 17(9): 1201-1214.
- Velayudhan J, Hughes NJ, McColm AA, Bagshaw J, Clayton CL, Andrews SC, Kelly DJ. Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. *Molecular Microbiology* 2000, 37 (2) : 274-286.

10. Anexos

The complete genome sequence of *Francisella noatunensis* subsp. *orientalis* strains FNO12, FNO24 and FNO190

Lucas Amorim Gonçalves¹, Siomar de Castro Soares², Felipe Luiz Pereira², Fernanda Alves Dorella², Alex Fiorini de Carvalho², Carlos Augusto Gomes Leal², Vasco Azevedo¹, Henrique César Pereira Figueiredo^{2*}

¹ Laboratory of Cellular and Molecular Genetics (LGCM) – Federal University of Minas Gerais, Belo Horizonte, Brazil.

² National Reference Laboratory for Aquatic Animal Diseases (AQUACEN), Ministry of Fisheries and Aquaculture – Federal University of Minas Gerais, Belo Horizonte, Brazil.

*Correspondence: Henrique César Pereira Figueiredo (figueiredoh@yahoo.com)

E-mail addresses:

Lucas Amorim Goncalves - amorimlucas@gmail.com

Siomar de Castro Soares - siomars@gmail.com

Felipe Luiz Pereira - felipe@flpsw.com.br

Fernanda Alves Dorella - fernandadorella@gmail.com

Alex Fiorini de Carvalho - alexficar@gmail.com

Carlos Augusto Gomes Leal - carlosleal@vet.ufmg.br

Vasco Ariston de Carvalho Azevedo - vasco@icb.ufmg.br

Henrique César Pereira Figueiredo - figueiredoh@yahoo.com

Keywords

Francisella noatunensis subsp. *Orientalis*, pathogen, genome sequencing

Abstract

The genus *Francisella* is composed of Gram-negative, pleomorphic, strictly aerobic and non-motile bacteria, which are capable of infecting a variety of terrestrial and aquatic animals. Among them, *Francisella noatunensis* subsp. *orientalis* (FNO) stands out as the causative agent of pyogranulomatous and granulomatous infections in fish. Accordingly, FNO is responsible for high mortality rates of freshwater fish, specially Tilapia, cultivated in the United States, United Kingdom, Japan, Taiwan, Jamaica, Costa Rica, Brazil and other Latin American countries. This scenario shows that more genomic data need to be obtained for this emergent pathogen to describe its virulence mechanisms. In the current study, we present the genome sequences of *Francisella noatunensis* subsp. *orientalis* strains FNO12, FNO24 and FNO190. The genomes include one

circular chromosome of 1,859,720 bp, 32% GC content, 1,538 proteins coding and 363 pseudo genes for FNO12; one circular chromosome of 1,862,322 bp, 32% GC content, 1,537 proteins coding and 365 pseudo genes for FNO24 and one circular chromosome of 1,859,595 bp, 32% GC content, 1,539 proteins coding and 362 pseudo genes for FNO190.

Introduction

Edward Francis (1872-1957), an American bacteriologist, described in 1922 the bacterium that causes tularemia in humans, *Francisella tularensis*. This bacterium is the most studied of its genus [1-2]. Until recently, the genus *Francisella* consisted of only two species: *F. tularensis* subsp. *tularensis* e *F. philomiragia* subsp. *philomiragia*. However, new species and new strains were isolated, such as *F. noatunensis* subsp. *noatunensis* e *F. noatunensis* subsp. *orientalis* (syn. *F. asiatica*) [1], being the latter recognized as one of the most important pathogens of cultured tilapia (*Oreochromis spp.*) [3].

F. noatunensis subsp. *orientalis* (FNO) is the etiologic agent of pyogranulomatous and granulomatous infections in fish. In the last few years, FNO has been responsible for a large number of deaths of tilapia and other freshwater species cultured in the United States, United Kingdom, Japan, Taiwan, Jamaica, Costa Rica, Brazil and other Latin American regions [4-5-6]. Nevertheless, besides infecting important cultivable species such as tilapia, threeline grunt (*Parapristipoma trilineatum*) and hybrid striped bass (*Morone chrysops* X *Morone saxatilis*), this bacterium is also capable of infecting wild fish such as guapote tigre (*Parachromis managuensis*) [4-5].

Although the disease caused by FNO has presented a high mortality rate during outbreaks and has been reported in several countries, the phylogenomic relationships among isolates from different countries and the evolutionary history of this pathogen are still poorly characterized. Therefore, the strains presented here were isolated from three different regions and outbreaks in order to characterize the genetic diversity of the microorganism *F. noatunensis* subsp. *orientalis* strains FNO12, FNO24 e FNO190.

Classification and features

This genus (Phylum: Proteobacteria; Class: Gammaproteobacteria; Order: Thiotrichales; Family: Francisellaceae) is characterized by strictly aerobic, non-motile, pleomorphic Gram-negative bacteria of 0.5-1.5 µm (**Table 1**). They are negative for nitrate reduction; adonitol, arabinose, cellobiose, esculin, galacturonate, glucuronate, malonate, mannitol, melibiose, raffinose, rhamnose, palatinose, and 5-ketogluconate fermentation; C14 lipase, cystine arylamidase, para-phenylalanine deaminase, tetrathionate reductase, trypsin, urease, valine arylamidase, α-chymotrypsin, α-fucosidase, α-galactosidase, α-mannosidase, and β-glucuronidase activity; and acid production from lactose. Additionally, they are positive for acid phosphatase, alkaline phosphatase, C4 and C8 esterase, lipase, naphtol-AS-BI-phosphohydrolase, and β-lactamase activity; and acid production from maltose [7].

Figure 1 shows a phylogenetic tree made based on 16S sequences. The result indicates a division into two groups: the first composed of the fish pathogens *F. noatunensis* subsp. *noatunensis* and *F. noatunensis* subsp. *orientalis*, plus *F. philomiragia*; and the second basically composed of *F. tularensis*. Additionally, in the first group, the strains that are the focus of this work were clustered within a subdivision with other two strains of *F. noatunensis* subsp. *orientalis*: Ehime-1 and Toba04.

Table 1. Classification and general features of *Francisella noatunensis* subsp. *orientalis* strains FNO12, FNO24 and FNO190 according to the MIGS recommendations[8]

MIGS ID	Property	Term	Evidence code*
	Classification	Domain <i>Bacteria</i>	TAS[7]
		Phylum <i>Proteobacteria</i>	TAS[7]
		Class <i>Gammaproteobacteria</i>	TAS[7]
		Order <i>Thiotrichales</i>	TAS[7]
		Family <i>Francisellaceae</i>	TAS[7]
		Genus <i>Francisella</i>	TAS[7]
		Species <i>Francisella noatunensis</i> subsp. <i>orientalis</i>	TAS[7]
		Type strain FNO12, FNO24 and FNO190	TAS
	Gram stain	Gram-negative	TAS[7]
	Cell shape	pleomorphic	TAS
	Motility	Non-motile	TAS[7]
	Sporulation	Not reported	TAS
	Temperature range	Mesophilic (15-34°C)	TAS[7]
	Optimum temperature	< 25°C	TAS[9]
	pH range; Optimum	Not reported	
	Salinity	Not reported	
MIGS-22	Oxygen requirement	Strictly aerobic	TAS[7]
	Carbon source	Not reported	TAS[7]
	Energy source	Not reported	
MIGS-6	Habitat	Host and aquatic environment	
	Salinity	Not reported	
MIGS-15	Biotic relationship	Intracellular facultative pathogen	TAS[7]
MIGS-14	Pathogenicity	Pathogenic	TAS[7,6]
MIGS-4	Geographic location	FNO12 – Brazil / State of Minas Gerais / Areado city FNO24 – Brazil / State of Minas Gerais / Alterosa city FNO190 – Brazil / State of São Paulo / Santa fé do Sul city	

MIGS-5	Sample collection time	FNO12 and FNO24 – 08/05/2012 FNO190 – 10/11/2013	
MIGS-4.1	Latitude	Not reported	
MIGS-4.2	Longitude	Not reported	
MIGS-4.4	Altitude	Not reported	

* Evidence codes - TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project[10].

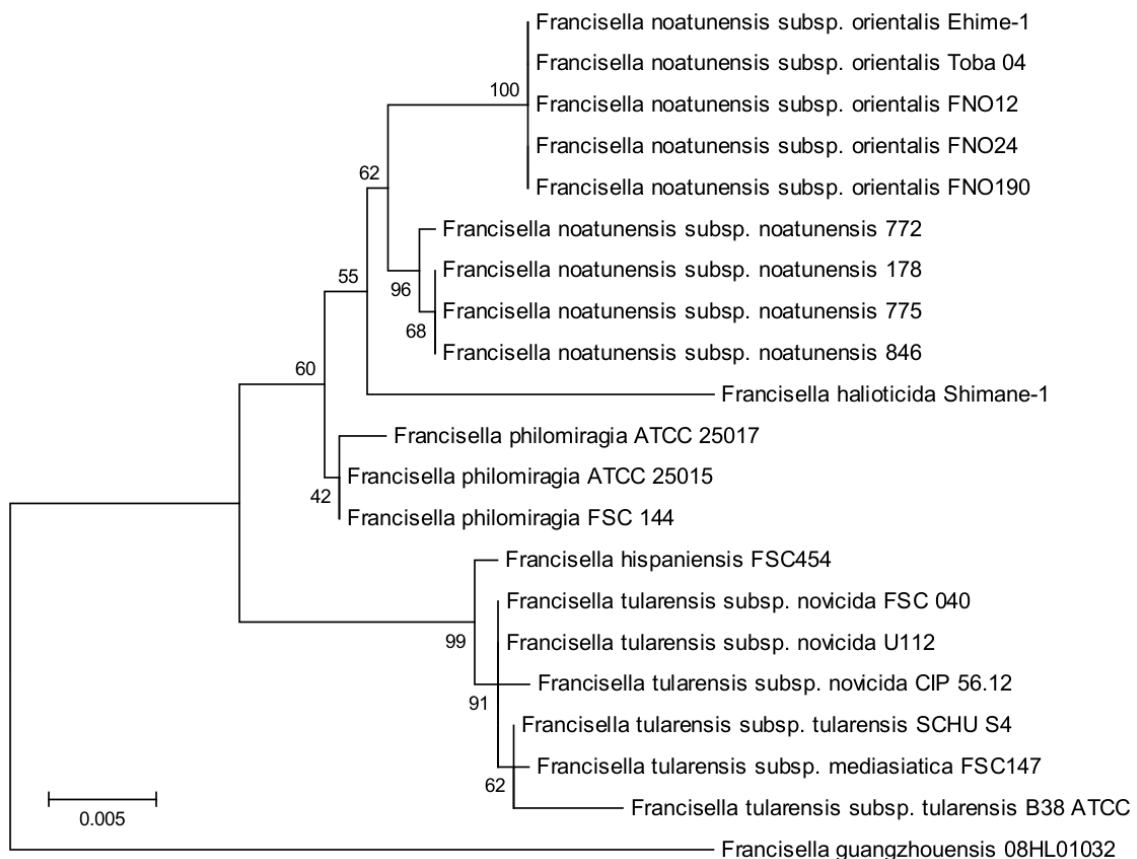


Figure 1. Phylogenetic tree of *F. noatunensis* subsp. *orientalis* strains FNO12, FNO24 and FNO190 representing their relative position in the genus *Francisella* based on 16S sequences. The statistical method used was Maximum Likelihood and the bootstrap number was 1000. Thus, the values next to the nodes represent the percentage on the number of times, in 1000 repetitions, in which that clade was formed.

Genome sequencing information

Genome project history

In the present study, it was determined the nucleotide sequence of *Francisella noatunensis* subsp. *orientalis* FNO12, FNO24 and FNO190 genomes, the first isolated from Nile tilapia kidney and the other two from spleen. Sequencing, assembly and annotation were performed by the Laboratory of Cellular and Molecular Genetics (LGCM) and the National Reference Laboratory for Aquatic Animal Diseases (AQUACEN) teams, both located at Federal University of Minas Gerais, Belo Horizonte, Minas Gerais, Brazil. Table 2 presents the project information and its association with MIGS version 2.0 compliance[8].

Table 2. Project information

MIGS ID	Property	FNO12	FNO24	FNO190
MIGS-31	Finishing quality	Finished	Finished	Finished
MIGS-28	Libraries used	Fragment	Fragment	Fragment
MIGS-29	Sequencing platforms	Ion Torrent PGM™ and Illumina	Ion Torrent PGM™	Ion Torrent PGM™
MIGS-31.2	Fold coverage	-	-	-
MIGS-30	Assemblers	Mira and Newbler	Mira and Newbler	Mira and Newbler
MIGS-32	Gene calling method	RAST (Rapid Annotations using Subsystems Technology)	RAST (Rapid Annotations using Subsystems Technology)	RAST (Rapid Annotations using Subsystems Technology)
	Locus Tag	FNO12	FNO24	FNO190
	Genbank ID	-	-	-
	Genome Database release	-	-	-
	GOLD ID	Gb0109929	Gb0109930	Gb0109931
	BIOPROJECT	PRJNA232116	PRJNA234502	PRJNA240882
	Project relevance	Pathogen	Pathogen	Pathogen
	Source Material Identifier	-	-	-

Growth conditions and DNA isolation

Francisella noatunensis subsp. *orientalis* FNO12, FNO24 and FNO190 were cultured in cysteine heart agar supplemented with 2% bovine hemoglobin (CHA; BD Biosciences, USA) and in Mueller-Hinton cation-adjusted broth supplemented with 2% VX (Laborclin, Brazil) and 1% glucose, and incubated at 28°C for 4-7 days [6]. Genomic DNA was extracted by the use of the Maxwell 16® Research Instrument (Promega, USA) according to the manufacturer's recommendations. Briefly, (i) the samples were lysed in the presence of a chaotropic agent and a detergent; (ii) nucleic acids were bound to silica magnetic particles; (iii) bound particles were washed, being isolated from other cell components, and (iv) nucleic acids were eluted into a formulation for sequencing. The extracted DNA was stored at -80°C until use.

Genome sequencing and assembly

The genome sequencing of the FNO12 strain was performed with the MiSeq platform (Illumina®) while the genome sequencing of the FNO24 and FNO190 strains. Fragment library was used for both platforms. The quality of the raw data was analyzed using FastQC [11] and the assembly was performed using the Mira 3.9 software [12] and Newbler 2.9 (Roche), being the *ab initio* strategy applied. The assemblies of FNO12, FNO24 and FNO190 produced a total of 15, 17 and 16 contigs, respectively. Regarding the coverage, the first strain resulted in a value of 1382,15x; the second, in a value of 79,82x; and the third, in a value of 203,43x. Additionally, the strains FNO12, FNO24 and FNO190 presented an N50 value of 275,043; 87,100; and 237,022; respectively. The gap filling was conducted through the use of CLC Genomics Workbench 7 (Quiagen, USA) after constructing scaffolds using Optical Map from OpGen as reference for FNO12 and the first strain as reference to FNO24 and FNO190.

Genome annotation

The automatic annotation was done using the RAST (Rapid Annotations using Subsystems Technology) software [13]. Additionally, tRNA and rRNA predictions were conducted using the tRNAscan-SE Search Server [available on: <http://lowelab.ucsc.edu/tRNAscan-SE/>] and the RNAmmer [(available on: <http://www.cbs.dtu.dk/services/RNAmmer/>)], respectively. It was also performed a manual curation of the annotation using the Artemis software [14] and the UniProt database (<http://www.uniprot.org>). All putative frameshifts were manually curated based on the coverage and CLC Genomics Workbench 7 (Quiagen, USA) software was used to correct indel errors in regions of homopolymers.

Genome properties

The genomes are composed of a circular chromosome with average sizes of 1860000 bp (Table 3). The G+C content in the three strains is 32% and the number of pseudogenes is relatively high, 363 on average. The FNO24 strain has a larger number of protein coding genes and one RNA-coding gene less than the other two strains. Additionally, for the FNO12 and FNO190 strains, it was annotated 1280 genes with function prediction whereas, for the strain FNO24, it was annotated 1282. However, that resulted in 621 hypothetical proteins in the three genomes (Table 4). Table 5 summarizes the number of genes associated with general COG functional categories and the figures 2, 3 and 4 show the disposition of RNAs and CDSs in forward and reverse strands plus the GC content and GC skew for all three strains discussed in this study.

Table 3. Summary of genome: one chromosome

Label	Size (bp)	Topology	Genbank ID	Strain
Chromosome	1859720	Circular	-	FNO12
Chromosome	1862322	Circular	-	FNO24
Chromosome	1859595	Circular	-	FNO190

Table 4. Genome statistics

Attribute	FNO12	FNO24	FNO190
Genome size (bp)	1859720	1862322	1859595
DNA coding (bp)	1348998	1343370	1350675
DNA G+C (bp)	445223	443266	445755
DNA scaffolds	1	NA?	NA?
Total genes	1951	1952	1951
Protein coding genes	1538	1537	1539
RNA genes	50	49	50
Pseudo genes	363	365	362
Genes in internal clusters	??		
Genes with function prediction	1280	1282	1280
Genes assigned to COGs	1327	1327	1326
Genes with Pfam domains	1562	1564	1561
Genes with signal peptides	128	128	126
Genes with transmembrane helices	531	531	534
CRISPR repeats	0	0	0

Table 5. Number of genes associated with general COG functional categories.

Code	FNO12	%age	FNO24	%age	FNO190	%age	Description
J	152	7,9957	152	7,9873	152	7,9957	Translation, ribosomal structure and biogenesis
A	1	0,0526	1	0,0525	1	0,0526	RNA processing and modification
K	47	2,4723	47	2,4697	47	2,4723	Transcription
L	74	3,8926	74	3,8885	74	3,8926	Replication, recombination and repair
B	0	0	0	0	0	0	Chromatin structure and dynamics
D	16	0,8416	16	0,8407	16	0,8416	Cell cycle control, Cell division, chromosome partitioning
V	17	0,8442	17	0,8933	17	0,8442	Defense mechanisms
T	16	0,8416	16	0,8407	16	0,8416	Signal transduction mechanisms
M	116	6,1020	116	6,0956	115	6,0494	Cell wall/membrane biogenesis
N	10	0,5260	10	0,5254	10	0,5260	Cell motility

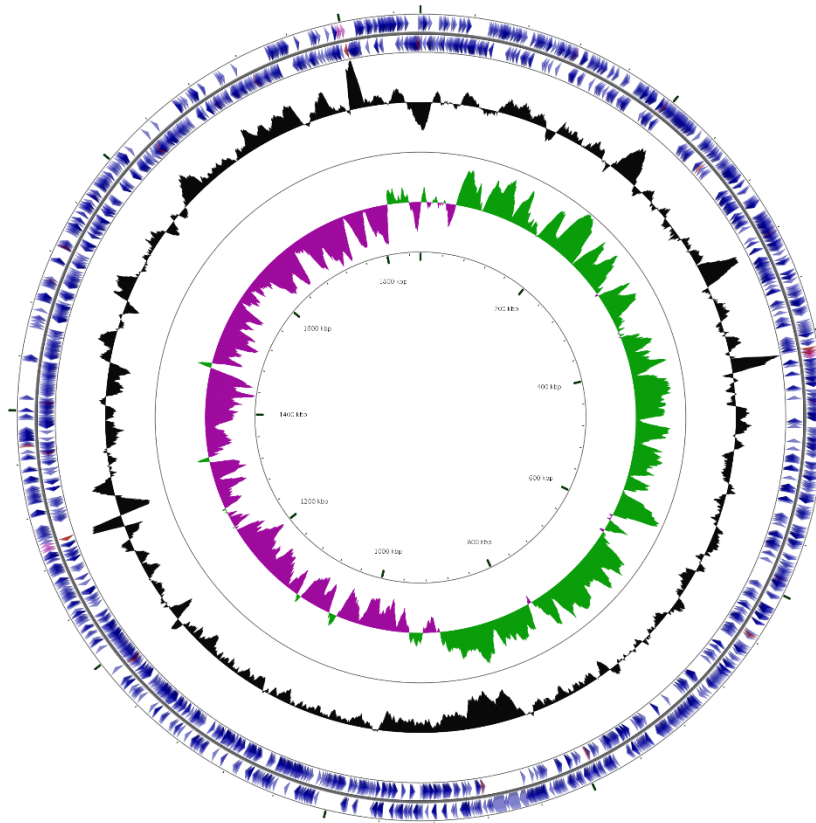
U	36	1,8937	36	1,8917	36	1,8937	Intracellular trafficking and secretion
O	68	3,5770	68	3,5733	68	3,5770	Posttranslational modification, protein turnover, chaperones
C	94	4,9447	94	4,9395	94	4,9447	Energy production and conversion
G	85	4,4713	85	4,4666	87	4,5765	Carbohydrate transport and metabolism
E	182	9,5739	182	9,5638	184	9,6791	Amino acid transport and metabolism
F	57	2,9984	57	2,9952	57	2,9984	Nucleotide transport and metabolism
H	80	4,2083	80	4,2038	80	4,2083	Coenzyme transport and metabolism
I	73	3,8400	73	3,8360	73	3,8400	Lipid transport and metabolism
P	74	3,8926	74	3,8885	76	3,9978	Inorganic ion transport and metabolism
Q	40	2,1041	40	2,1019	40	2,1041	Secondary metabolites biosynthesis, transport and catabolism
R	173	9,1004	173	9,0909	174	9,1530	General function prediction only
S	99	5,2077	99	5,2023	98	5,1551	Function unknown
-	574	30,1946	576	30,2679	575	30,2472	Not in COGs

^aThe percentage is based on the total number of protein coding genes in the annotated genome.

^bThe total not correspond to quantidade final de CDSs de cada genoma, because some genes are associated with more than one COG functional categories.

Accession: unknown
Length: 1,859,720 bp

- CDS
- rRNA
- tRNA
- Other
- GC content
- GC skew+
- GC skew-

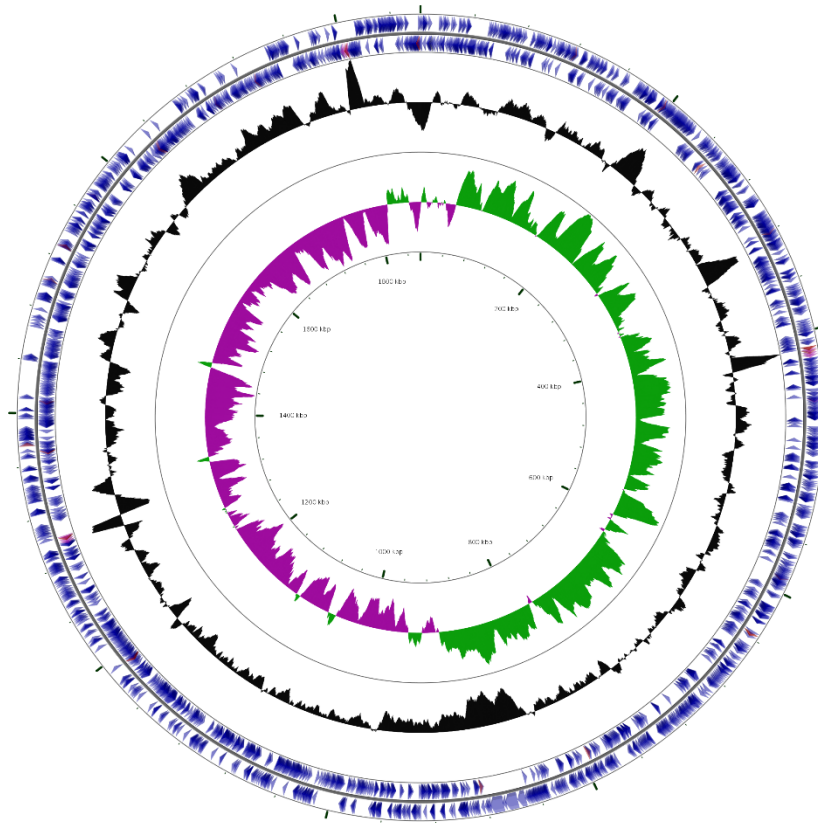


Francisella noatunensis subsp. *orientalis* FNO12

Figure 2. Graphical circular map of the genome. From outside to the center: CDSs and on forward strand; CDSs and RNAs on reverse strand; GC content; and, GC skew.

Accession: unknown
Length: 1,862,322 bp

- CDS
- tRNA
- rRNA
- Other
- GC content
- GC skew+
- GC skew-

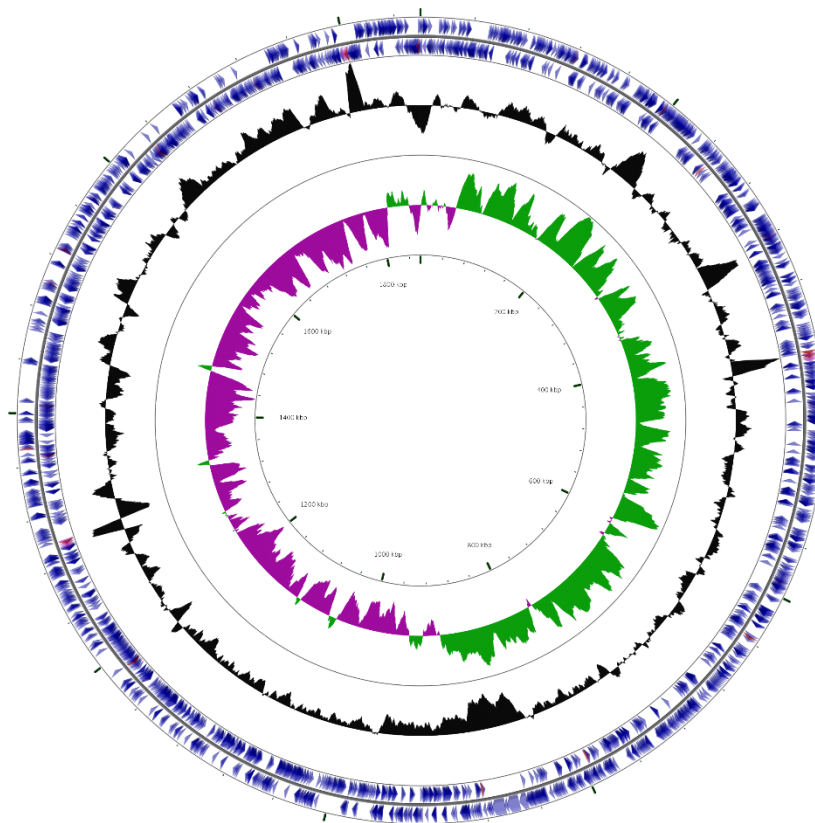


Francisella noatunensis subsp. *orientalis* FNO24

Figure 3. Graphical circular map of the genome. From outside to the center: CDSs and on forward strand; CDSs and RNAs on reverse strand; GC content; and, GC skew.

Accession: unknown
Length: 1,859,595 bp

CDS
tRNA
rRNA
Other
GC content
GC skew+
GC skew-



Francisella noatunensis subsp orientalis FNO190

Figure 4. Graphical circular map of the genome. From outside to the center: CDSs and on forward strand; CDSs and RNAs on reverse strand; GC content; and, GC skew.

Conclusions

Here, it was presented three genomes of an important fish pathogen that, despite being isolated from different outbreaks, are very similar if considered the brief analysis of this work. All tables and figures suggest the clonality of the strains with minor differences quantity of pseudogenes, and number of CDSs, RNAs and protein coding genes. Furthermore, FNO12, FNO24, FNO190, *F. noatunensis* subsp. *orientalis* Ehime-1 and *F. noatunensis* subsp. *orientalis* Toba04 formed the same clade 100% of times in 1000 repetitions of bootstrap, which shows the great similarity among these microorganisms.

Acknowledgement

This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Ministério da Pesca e Aquicultura and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). We also acknowledge support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

References

1. Sridhar S, Sharma A, Kongshaug H, Nilsen F, Jonassen I. Whole genome sequencing of the fish pathogen *Francisella noatunensis* subsp. *orientalis* Toba04 gives novel insights into *Francisella* evolution and pathogenicity. *BMC Genomics* 2012; 13:598.
2. Soto E, Kidd S, Mendez S, Marancik D, Revan F, Hiltchie D, Camus A. *Francisella noatunensis* subsp. *orientalis* pathogenesis analyzed by experimental immersion challenge in Nile tilapia, *Oreochromis niloticus* (L.). *Vet Microbiol* 2013; 164(1-2):77-84.
3. Soto E, Revan F. Culturability and persistence of *Francisella noatunensis* subsp. *orientalis* (syn. *Francisella asiática*) in sea-and freshwater microcosms. *Micro Ecol* 2012; 63(2):398-404.
4. Soto E, Illanes O, Hilchie D, Morales Ja, Sunyakumthorn P, Hawke JP, Goodwin AE, Riggs A, Yanong RP, Pouder DB, Francis-Floyd R, Arauz M, Bogdanovic L, Castillo-Alcala. Molecular and immunohistochemical diagnosis of *Francisella noatunensis* subsp. *orientalis* from formalin-fixed, paraffin-embedded tissues. *J Vet Diagn Invest* 2012; 24(5):840-5.
5. Soto E, Abrams SB, Revan F. Effects of temperature and salt concentration on *Francisella noatunensis* subsp. *Orientalis* infections in Nile tilapia *Oreochromis niloticus*. *Dis Aquat Organ* 2012; 101(3):217-23.
6. Leal CAG, Tavares GC, Figueiredo HCP. Outbreaks and genetic diversity of *Francisella noatunensis* subsp. *orientalis* isolated from-raised Nile tilapia (*Oreochromis niloticus*) in Brazil. *GMR* 13 (3): 5704-5712 (2014)
7. Birkbeck TH, Feist SW, Verner-Jeffreys DW. *Francisella* infections in fish and shellfish. *J Fish Dis* 2011; 34(3):173-87.
8. Field D, Garrity G, Gray T, *et al.* The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* 2008;26:541–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2409278&tool=pmcentrez&rendertype=abstract>. Accessed September 2, 2014.
9. Ottem, K.F., Nylund, A., Karlsbakk, E., Friis-Møller, A. and Kamaishi, T. Elevation of *Francisella philomiragia* subsp. *noatunensis* to *Francisella noatunensis* comb. nov. [syn. *Francisella piscicida* Ottem et al. (2008) syn. nov.] and characterization of *Francisella noatunensis* subsp. *orientalis* subsp. nov., two important fish pathogens. *J of App. Mic.* 2009; 106: 1231–1243
10. Anon. Gene Ontology Consortium-Experimental Evidence Codes. Available at: <http://geneontology.org/page/experimental-evidence-codes>.

11. Anon. Babraham Bioinformatics - FastQC. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
12. Chevreux, B; Wetter, T; Suhai S. Genome Sequence Assembly Using Trace Signals and Additional Sequence Information. *Comput. Sci. Biol. Proc. Ger. Conf. Bioinforma.* 1999;**99**:45–56.
13. Aziz RK, Bartels D, Best AA, *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 2008;**9**:75. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2265698&tool=pmcentrez&rendertype=abstract>. Accessed July 9, 2014.
14. Rutherford K, Parkhill J, Crook J, *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* 2000;**16**:944–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11120685>. Accessed August 17, 2014.