

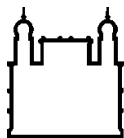
MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ

Mestrado em Programa de Biologia Computacional e Sistemas

EPIDEMIOLOGIA GENÔMICA DE BORDETELLA PERTUSSIS NO BRASIL

DIEGO DUQUE CAMBUY

Rio de Janeiro  
Julho de 2014



Ministério da Saúde

**FIOCRUZ**  
**Fundação Oswaldo Cruz**

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

*Diego Duque Cambuy*

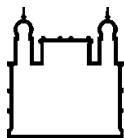
Epidemiologia Genômica de *Bordetella pertussis* no Brasil

Tese ou dissertação apresentada ao Instituto  
Oswaldo Cruz como parte dos requisitos para  
obtenção do título de Mestre em Ciências

**Orientador (es):** Prof. Dra. Ana Carolina Paulo Vicente

**RIO DE JANEIRO**

Julho de 2014



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

***AUTOR: DIEGO DUQUE CAMBUY***

**Epidemiologia Genômica de *Bordetella pertussis* no Brasil**

**ORIENTADOR (ES): Prof. Dra. Ana Carolina Paulo Vicente**

**Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_**

**EXAMINADORES:**

**Prof. Dr. Marcos Paulo Catanho de Souza - Presidente** (Fundação Oswaldo Cruz)

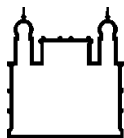
**Prof. Dr. Antônio Basílio de Miranda** (Fundação Oswaldo Cruz)

**Prof. Dr. Cristiane Carneiro Thompson** (Universidade Federal do Rio de Janeiro)

**Prof. Dr. Fábio Passetti** (Instituto Nacional do Câncer)

**Prof. Dr. Michel Francisco Abanto Marin** (Fundação Oswaldo Cruz)

Rio de Janeiro, de 31 de Julho 2014



Ministério da Saúde

FIOCRUZ

**Fundação Oswaldo Cruz**

**Anexar a cópia da Ata que será entregue pela SEAC já assinada.**

## **AGRADECIMENTOS**

À minha orientadora, Ana Carolina, pela oportunidade do desenvolvimento desse trabalho;

À Fernanda Jannechevitz pelo apoio em todos os momentos;

À Fernanda Jannechevitz e Érica Miranda pela ajuda do trabalho experimental;

Ao Professor Marcos Catanho pela revisão deste manuscrito;

À CAPES e FIOCRUZ pelo apoio financeiro.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

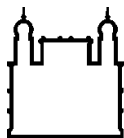
### Epidemiologia Genômica de *Bordetella pertussis* no Brasil

#### RESUMO

#### DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Diego Duque Cambuy

A coqueluche, ou pertússis, é uma doença do trato respiratório causada principalmente pela bactéria *Bordetella pertussis*. Após 50 anos de vacinação, pertussis reemergiu, passando a ser a doença imunoprevenível mais frequente mesmo em países desenvolvidos. Várias são as hipóteses para a re-emergência de pertússis, uma delas é a adaptação do patógeno frente à vacinação. Linhagens contemporâneas de *B. pertussis* diferem de linhagens do período pré-vacinal, especialmente em genes codificadores de proteínas usadas na produção de vacinas acelular. Esta re-emergência também tem sido observada no Brasil, assim, realizamos a caracterização genética por MLST baseado nesses genes, de 26 isolados *B. pertussis* de surtos de três regiões brasileiras (Norte, Sul e Nordeste). Foram identificados dois perfis alélicos, em 24 isolados: *prn2-ptxS1A-fim3B-ptxP3*, de surtos (2008-2013) de Alagoas, Pernambuco e Rio Grande do Sul - e o perfil *prn2-ptxS1A-fim3A-ptxP3*, em dois isolados de Pará/2004. Análises filogenéticas agruparam esses perfis com isolados do período pós vacinal de outras partes do globo. Deste conjunto, três do perfil mais frequente e um do perfil menos frequente, tiveram seus genomas sequenciados na plataforma GS 454 Junior. A comparação desses genomas com outros genomas de *B. pertussis* disponíveis em dados públicos não identificou SNPs ou genes únicos que caracterizassem os isolados do Brasil. Este estudo desenvolveu uma metodologia que permitiu definir a posição da IS481 nos genomas, e uma delas corresponde a um gene relacionado a regulação da transcrição da família MarR, Análise filogenômica, baseada em 826 SNPs, demonstrou que os isolados recentes do Brasil da linhagem pandêmica que presente em todos os continentes, exceto a África. Foi observado também que as relações filogenéticas inferidas pelo MLST são semelhantes àquelas inferidas quando se utiliza o genoma completo, isso denota a pressão seletiva sobre esses genes. Sendo assim, a cepa utilizada na produção da vacina no Brasil, que apresenta o perfil alélico *prn1-ptxS1D - fim3A-ptxP2*, pode não ser capaz de gerar uma resposta imune protetora frente às linhagens circulantes no país. Este estudo traz, pela primeira vez, informações genéticas e genômicas de isolados de *B. pertussis* do Brasil, país que apresenta cobertura vacinal bastante heterogênea, que utiliza, oficialmente, a vacina celular, mas que, também, aplica a vacina acelular. As informações reveladas neste estudo podem auxiliar a tomada de ações para o controle de pertússis no Brasil, além do conhecimento sobre epidemiologia e evolução de *B. pertussis*.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

### Genomic Epidemiology of *Bordetella pertussis* in Brazil

#### ABSTRACT

#### MASTER DISSERTATION IN COMPUTATIONAL BIOLOGY SYSTEMS

Diego Duque Cambuy

Pertussis more commonly referred as whooping cough is respiratory tract disease mainly caused by the bacteria *B. pertussis*. After 50 years of vaccination pertussis reemerged, becoming the most frequent vaccine preventable disease in developed countries. Many hypotheses have been proposed for the re-emergence of pertussis, one being the pathogen adaptation in a vaccinated environment. Current pertussis strains are different than those from the prevaccination era, especially in genes that code for proteins used in acellular pertussis vaccines. This re-emergence is also observed in Brazil, therefore we characterized 26 isolates from 3 regions of Brazil (North, South, Northeast) using an MLST approach based on these genes. We identified two allelic profiles, 24 isolates from the states of Rio Grande do Sul (2008-2009), Alagoas (2008-2009), Pernambuco (2013) and Pará (2004) presented the prn2-ptxS1A-fim3B-ptxP3 allelic profile, while 2 isolates from Pará (2004) presented the prn2-ptxS1A-fim3A-ptxP3 allelic profile. Phylogenetic analysis branch these two allelic profiles along with other post vaccination isolates around the globe. Four isolates, three from the dominant profile and one from the less frequent profile, had their genomes completed sequenced on the GS 454 Junior Platform. We compared these genomes with others available in public databases and no SNP or unique genes were identified in the Brazilian genomes. This study also developed a methodology that identifies the location of the repetitive region IS481, and what genes it interrupted. One of them was the MarR transcriptional regulator gene. Phylogenomic analysis based on 826 SNPs revealed that Brazilian *B. pertussis* lineages are part of the current pandemic lineage present in all continents, except Africa. We also observed that phylogenomic relationships are similar to MLST's. Therefore, strain used for pertussis vaccine in Brazil, that presents the prn1-ptxS1D -fim3A-ptxP2 allelic profile, might not be able to induce immune response to the current lineage circulating in the country. This is the first study with genetic and genomic informations of *B. pertussis* isolates in Brazil, which is a country with heterogeneous vaccine coverage and mixed and has both cellular and acellular vaccine administrated to the population. Information brought with this study can help the decision making on the control of pertussis in Brazil and gives new insights on the epidemiology and evolution of *B. pertussis*.



## Índice

1. Introdução .....	2
1.1. Pertússis .....	2
1.2. Aspectos Históricos da pertússis .....	2
1.3. Epidemiologia .....	4
1.4. <i>B. pertussis</i> : Genética e Determinantes de Virulência.....	8
1.5. Histórico da genômica de <i>B. pertussis</i> .....	13
1.6. Papel da IS481 em <i>B. pertussis</i> .....	16
1.7. Justificativa.....	17
2. Objetivos .....	19
2.1. Objetivo geral .....	19
2.2. Objetivos específicos .....	19
3. Metodologia .....	21
3.1. Trabalho Experimental.....	23
3.1.1. Isolados de <i>B. pertussis</i> .....	23
3.1.2. Extração de DNA, amplificação e sequenciamento dos genes para o MLST 24	
3.2. Análise de sequências de multilocus (MLST – Multilocus Sequence Analysis) 26	
3.3. Sequenciamento do genoma utilizando a plataforma GS 454 Junior .....	27
3.5. Identificação de SNPs .....	31
3.6. Sintenia.....	31
3.7. Identificação de regiões do genoma interrompidos pela IS481 .....	32
4. Resultados .....	38
4.1. Caracterização genética de isolados de <i>B. pertussis</i> do Brasil .....	38
4.1.1. MLST de genes de virulência, e genes de resistência .....	38
4.1.2. Relação genética entre isolados de <i>B. pertussis</i> .....	40
4.2. Genomas de <i>B. pertussis</i> do Brasil.....	41
4.2.1. Métricas do sequenciamento genômico por alto desempenho .....	41
4.2.2. Características dos genomas sequenciados e anotados .....	45
4.2.3. Identificação de genes ortólogos, genoma core e pangenoma de <i>B. pertussis</i> 46	
4.3. Filogenômica e SNPs.....	51
4.4. Sintenia.....	53
4.5. Identificação de genes interrompidos pela IS481 .....	56
5. Discussão .....	59
7. Referências Bibliográficas .....	72
8. Anexos .....	82

## Índice de Figuras

Figura 1. Linha do tempo com os aspectos históricos da pertussis.....	3
Figura 2. Mapa da re-emergência de pertussis.....	5
Figura 3. Gráfico com o número de casos de pertússis nos anos de 2008 a 2011.....	6
Figura 4. Distribuição de diferentes classes de genes ao longo do genoma de <i>B. pertussis</i> .....	11
Figura 5. Alinhamento múltiplo mostrando o gene <i>prn</i> .....	13
Figura 6. Representação gráfica do número de genomas de <i>B. pertussis</i> sequenciados desde o primeiro em 2003 .....	14
Figura 7. Fluxograma com a metodologia aplicada neste estudo.....	22
Figura 8. Representação parcial do alinhamento de todas as seqüências de Inserção IS481 de Tohama I .....	34
Figura 9. Representação da busca por regiões interrompidas pela IS481 .....	35
Figura 10. Distribuição dos pefis de MLST entre o conjunto isolados de <i>B. pertussis</i> .....	38
Figura 11. Relações genéticas entre <i>B. pertussis</i> a partir das sequencias dos genes .....	40
Figura 12. <i>Boxplot</i> com a qualidade dos <i>reads</i> em relação a posição da base ao longo do <i>read</i> . .....	43
Figura 13 . Gráfico com contagem do tamanho dos <i>reads</i> .....	44
Figura 14. Representação do genoma <i>core</i> e pangenoma de <i>B. pertussis</i> a partir 17 genomas representativos no tempo-espaco.....	47
Figura 15 : Diagrama de Venn mostrando o compartilhamento de genes entre os 4 genomas gerados neste estudo e o genoma referência Tohama I.....	50
Figura 16. Relação filogenética entre <i>B. pertussis</i> considerando o conjunto de 2050 genes ortólogos.....	52
Figura 17. Árvore de máxima verossimilhança construída com 826 SNPs.....	53
Figura 18. Representação da sintenia entre os 4 genomas dos 4 isolados brasileiros gerados neste estudo.....	55

## Lista de Tabelas

<b>Tabela 1.. Número de casos por Unidade da Federação em 2011.....</b>	<b>7</b>
<b>Tabela 2. Isolados De <i>B. pertussis</i> usados neste estudo sua localidade e ano de isolamento.....</b>	<b>.23</b>
<b>Tabela 3. Iniciadores para amplificação e/ou sequenciamento de genes de virulência para caracterização genética.....</b>	<b>..26</b>
<b>Tabela 4. Genomas usados para as análises deste estudo .....</b>	<b>28</b>
<b>Tabela 5. Isolados de <i>B. pertussis</i>, sua origem e perfil alélico dos genes de virulência.....</b>	<b>30</b>
<b>Tabela 6. Métrica do sequenciamento e montagem dos quatro genomas de isolados de <i>B. pertussis</i> do Brasil.....</b>	<b>39</b>
<b>Tabela 7. Características dos genomas sequenciados neste estudo ...</b>	<b>42</b>
<b>Tabela 8. Genes únicos dos genomas de isolados do Brasil em relação a Tohama I .....</b>	<b>49</b>
<b>Tabela 9. Posições interrompidas no genoma dos isolados do Brasil em relação ao genoma de referências Tohama I.....</b>	<b>57</b>

## Lista de abreviaturas

**ACT:** Toxina adenilato ciclase / *Arthemis Comparison Tool*

**AL:** Alagoas

**ATCC:** American-type Culture Colection

**BLAST:** *Basic Alignment Search Tool*

**CBMA:** Coleção de Bactérias da Meio Ambiente e Saúde

**DNA:** Ácido desoxirribonucleico

**DTP:** Difteria Tétano Pertússis

**FHA:** Filamentos de Hemoaglutinina

**Fim:** Fímbria

**GC:** Guanina-citosina

**IS:** Sequência de Inserção (*Insetion Sequence*)

**Kb:** Kilobases

**LGMM:** Laboratório de Genética Molecular de Micro-organismos

**Mb:** Megabases

**MLST:** *Multilocus Sequecing Typing*

**NBIC:** Centro de Bioinformática da Holanda (*Netherlands Bioinformatics Centre*)

**PA:** Pará

**PCR:** Reação em Cadeia da Polimerase (*Polmymerase Chain Reaction*)

**Pb:** Pares de bases

**PE:** Pernambuco

**PNI:** Programa Nacional de Imunização

**Prn:** Pertactina

**PTX:** Toxina Pertússis

**RS:** Rio Grande do Sul

**RAST:** *Rapid Annotation using Subsystems Technology*

**SNP:** Polimorfismo de básica única (*Single Nucleotide Polymorphism*)

**TCF:** Citotoxina traqueal

# Introdução

## **1. Introdução**

### **1.1. Pertússis**

Pertússis, coqueluche, ou tosse comprida, é uma doença do trato respiratório, altamente contagiosa. O quadro clínico apresenta três estágios: catarral, no qual o principal sintoma é a coriza; paroximal, caracterizado por crises de tosse e dificuldade de respirar e, por fim, o estágio de convalescência, quando a frequência da tosse diminui e inicia-se a recuperação (MATTOO; CHERRY, 2005).

### **1.2. Aspectos Históricos da pertússis**

A história de pertússis ao comparada com outras doenças epidêmicas como tuberculose, hanseníase, varíola, é relativamente recente. A primeira descrição de pertússis é de 1640, que relatava uma epidemia ocorrida na França, em 1578 (CONE, 1970) (Figura 1). Já foi proposto que a origem da *B. pertussis* seria a África ou a Índia, e que tenha sido introduzida na Europa em 1414(HEGERLE; GUIISO, 2013), porém esta é ainda uma hipótese.

O principal agente etiológico da pertússis é a bactéria *Bordetella pertussis*, isolada por Jules Bordet e Octave Gengou em 1906, quando foi denominada *Hamophilus pertussis* (BORDET; OCTAVE, 1906). Já no final da década de 20, o médico Thorvald Madsen começava os primeiros testes com uma vacina celular, composta por uma suspensão de *B. pertussis*. Com esses testes, realizados na Dinamarca, foi provado que pertussis seria imunoprevenível (BAKER; KATZ, 2004) e, no final dos anos 30, Pearl Kendrick combinou a vacina celular para pertussis juntamente com a vacina para diphteria e tetátano, dando origem à vacina DTP (Diphteria, Tétano, Pertússis) (KENDRICK; ELDERING; HEALTH, 1939). Esta vacina passou a ser aplicada em programas de vacinação a partir dos anos 50, fazendo com que o número de casos de pertússis, uma das maiores causas de morbidade e mortalidade entre crianças em todo o mundo, diminuísse significativamente (EDWARDS; HALASA, 2004).

A vacina celular está associada a reações adversas, desde irritação local até febre e convulsões, o que levou a uma resistência da população à vacinação, principalmente depois que surtos e epidemias deixaram de ser frequentes, perdurando até os dias atuais (ALLEN, 2013; LINNEMANN et al., 1975). Como consequência, cientistas partiram em busca de uma vacina, menos reatogênica, tendo sido desenvolvida uma vacina acelular.

A vacina acelular é composta por diferentes fatores de virulência de *B. pertussis*: pertactina (Prn), filamentos de hemoaglutinina (FHA), fímbrias do sorogrupo 1 e 2 (Fim1 e Fim2), e o principal fator de virulência de *B. pertussis*, a toxina pertussis (PTX). Com exceção da toxina pertussis, presentes em todas as versões de vacina acelular, a presença dos demais componentes é variável entre as diferentes vacinas (MOOI; VAN DER MAAS; DE MELKER, 2013a). Introduzida primeiramente no Japão, em 1981, essas vacinas substituíram gradualmente a vacina celular em vários países desenvolvidos (SATO; KIMURA; FUKUMI, 1984).

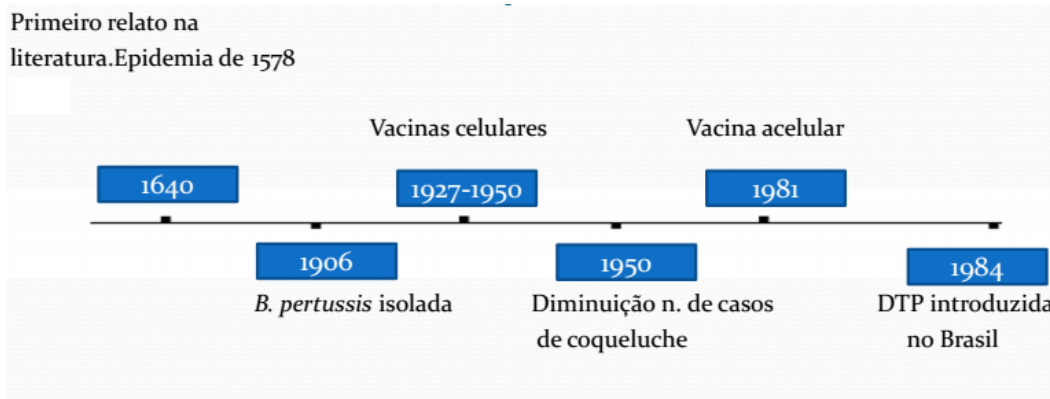
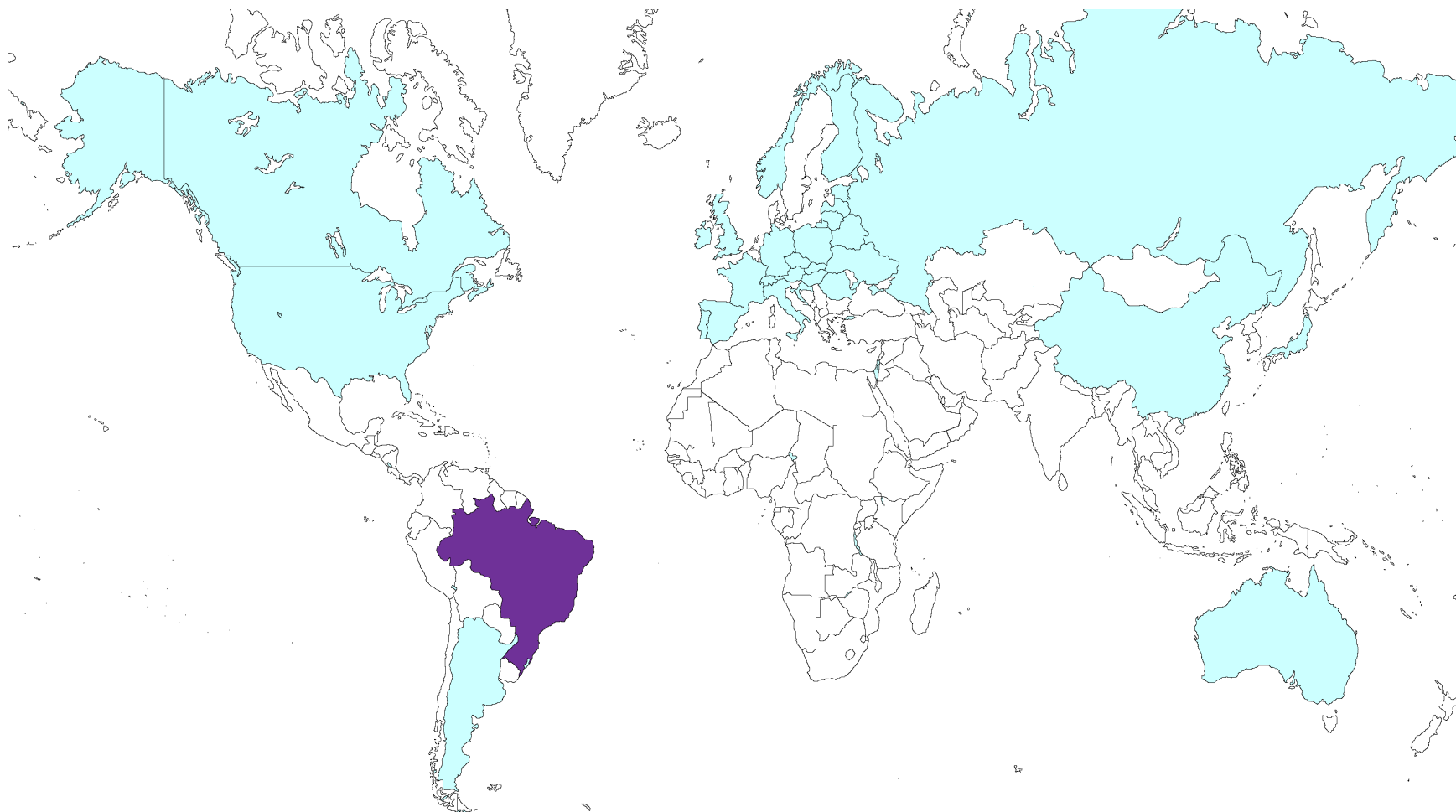


Figura 1. Linha do tempo com os aspectos históricos da pertússis



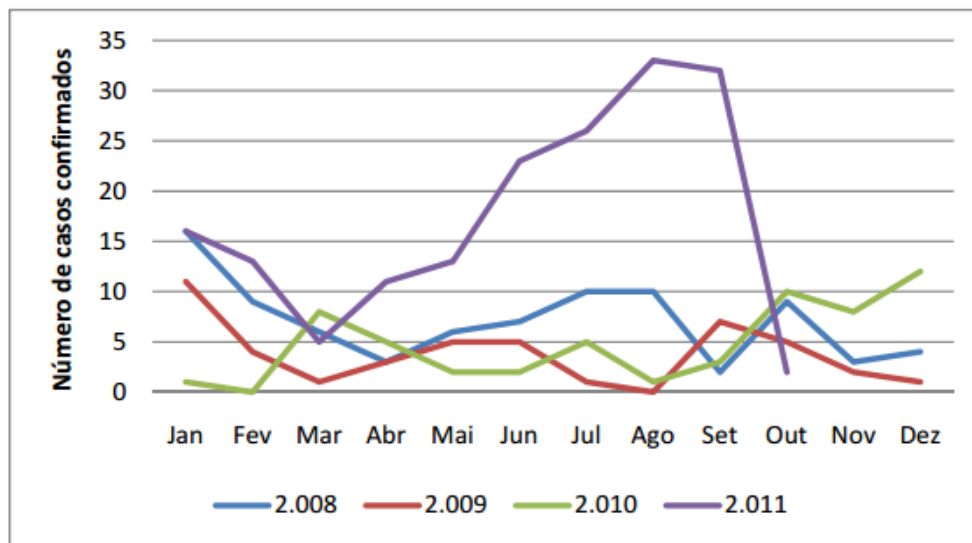
### **1.3 Epidemiologia**

De acordo com a Organização Mundial de Saúde, em 2008, foram relatados 195.000 casos de pertússis, sendo 95% destes em países em desenvolvimento (HEGERLE; GUIISO, 2013). Pertússis, que já foi a infecção mais frequente em crianças de até um ano de idade, passou a ser controlada após a introdução da vacina celular, virtualmente desaparecendo em países industrializados nos anos 70 e 80 (TARANGER et al., 2001). Porém, após mais de 50 anos, a infecção re-emergiu e, desde o início dos anos 90, um crescente número de casos vem sendo notificados, especialmente, em países com alta cobertura vacinal como Argentina, Austrália, Canadá, Estados Unidos, China e vários países da Europa. (CAMPINS et al., 2013; HALLANDER et al., 2005; PARADOWSKA-STANKIEWICZ; RUDOWSKA, 2013; TAN; TRINDADE; SKOWRONSKI, 2005). A re-emergência de pertússis também é observada no Brasil (RESP, 2011) (Figura 2)



**Figura 2. Mapa da re-emergência de pertússis no mundo.** Em azul estão destacados os países onde o número de casos de pertússis aumentou nos últimos anos. O Brasil demarcado em roxo também apresentou aumento do número de casos nos últimos anos (13-17).

No Brasil, pertússis é uma doença de notificação compulsória e a vacina faz parte do Programa Nacional de Imunização (PNI). A vacina celular é combinada com a vacina para o tétano e difteria e recentemente foi combinada com a vacina para o *Hamophilus influenzae* tipo B. A vacina acelular está disponível para adolescentes e adultos em clínicas particulares. Assim como em outros países, a vacinação ajudou a reduzir o número de casos, porém nos últimos anos o número de casos e surtos vem ocorrendo em vários estados do país. Os estados com o maior número de casos notificados são São Paulo, Rio Grande do Sul e Bahia (RESP, 2011). (Figura 3/ tabela 1)



**Figura 3. Gráfico com o número de casos de pertússis nos anos de 2008 a 2011.**  
Reprodução SINANnet/DDTR/CVE/CCD/SES-SP. (RESP, 2011)

**Tabela 1. Número de casos por Unidade da Federação em 2011.** Reprodução SINANnet/DDTR/CVE/CCD/SES-SP. (RESP, 2011)

UF	2011	
	Caso	Coefficiente de Incidência
RO	17	1,1
AC	9	1,3
AM	13	0,4
RR	3	0,7
PA	12	0,2
AP	1	0,2
TO	0	0,0
MA	11	0,2
PI	0	0,0
CE	4	0,0
RN	10	0,3
PB	1	0,0
PE	17	0,2
AL	23	0,7
SE	2	0,1
BA	65	0,4
MG	25	0,1
ES	4	0,1
RJ	30	0,2
SP	219	0,5
PR	32	0,3
SC	9	0,1
RS	53	0,5
MS	4	0,2
MT	0	0,0
GO	2	0,0
DF	17	0,7
<b>Total</b>	<b>583</b>	<b>0,3</b>

Muitas causas foram propostas para a re-emergência de pertússis em países desenvolvidos: aumento da vigilância; melhorias em técnicas de diagnóstico; substituição da vacina celular por uma vacina acelular e adaptação do patógeno frente à vacinação (CHIAPPINI et al., 2013). De fato, todos esses fatores podem ter contribuído para o aumento do número de casos e ou re-emergência da pertússis em todo o mundo. Em alguns países, o diagnóstico passou a ser definido não só pelo

isolamento da bactéria, mas também pela detecção de gene(s) da bactéria, o que pode ter aumentado o número de resultados positivos. No Brasil, o isolamento da bactéria continua sendo o padrão ouro para o diagnóstico (TREVIZAN; COUTINHO, 2008). A substituição da vacina celular por uma vacina acelular poderia ser outro fator já que a proteção conferida por ela é limitada, o que pode aumentar o número de indivíduos assintomáticos, representando um risco para os não vacinados (WARFEL; ZIMMERMAN; MERKEL, 2014).

Evidências sobre a adaptação do patógeno frente às vacinas vêm sendo demonstradas ao longo dos últimos anos. Vários estudos demonstram que as atuais linhagens causadoras de pertússis na Europa, Estados Unidos e Austrália são distintas daquelas da era pré vacinal, e são caracterizadas pela presença de polimorfismos nos genes codificadores de proteínas usadas nas vacinas acelulares (CASSIDAY et al., 2000; HALLANDER et al., 2007; OCTAVIA et al., 2012), além de apresentarem uma elevada expressão de genes de virulência (DE GOUW et al., 2014; KING et al., 2013).

Recente estudo de epidemiologia genômica mostra que esta é uma linhagem pandêmica que ainda não foi identificada na África, onde a vacinação ainda é recente (BART et al., 2014).

#### **1.4. *B. pertussis*: Genética e Determinantes de Virulência**

O gênero *Bordetella* compreende um grupo de nove espécies de bactérias aeróbicas restritas, com exceção de *B. petrii*, gram-negativas e de morfologia cocobacilar, que possuem a capacidade de colonizar o trato respiratório de animais e, em muitos casos, ocasionar doenças. São pequenas, com tamanho variando de 0.2-0.7 µm (REVIEWS, 2012). Por apresentarem crescimento lento e serem fastidiosas, o cultivo tende a ser difícil. Dentre as espécies do gênero, quatro delas *B. pertussis*, *Bordetella bronchiseptica*, *Bordetella parapertussis* e *Bordetella holmesii* são patógenos humanos causando infecções respiratórias. Enquanto as outras espécies do gênero são patógenos oportunistas, podendo ser isolados tanto do trato respiratório quanto de outras áreas (MOOI; HE; GUIZO, 2007; PARK et al., 2012a).

As espécies *B. pertussis*, *B. bronchiseptica* e *B. parapertussis* são relacionadas e por isso são chamadas de *Bordetellas* clássicas. Porém, *B. pertussis* infecta

exclusivamente humanos, enquanto que *B. parapertussis* é encontrada em ovinos e humanos, já *B. bronchiseptica* pode ser encontrada em uma variedade de mamíferos (VAN DER ZEE et al., 1996). A infecção por um desses parasitas pode levar a pertússis ou doenças similares, sendo *B.pertussis* o principal agente etiológico da mesma em humanos.

Análises por hibridação de DNA (CUMMINGS et al., 2004), sequenciamento de genes do metabolismo comum (MUSSER et al., 1986), e mais recentemente, genômica comparativa (PARKHILL et al., 2003), mostram que tanto *B. pertussis* quanto *B. parapertussis* evoluíram a partir de um ancestral comum com *B. bronchiseptica* através de massiva perda e ou inativação gênica, o que possivelmente determinou a especificidade pelos distintos hospedeiros atuais.

Apesar de já ter sido demonstrado em laboratório que *B. pertussis* é capaz de receber e manter plasmídeos, a ocorrência natural só foi relatada uma única vez (KAMACHI et al., 2006). Eventos de transferência lateral de genes também não são comuns, tendo sido relatados poucas vezes na literatura (DIAVATOPOULOS et al., 2006; PARK et al., 2012b). Parkhill e colaboradores (2003) sugerem que a raridade de eventos de transferência lateral de genes se deve ao fato da ausência de fagos e plasmídeos na espécie (PARKHILL et al., 2003). Sendo assim o principal mecanismo de variabilidade genética em *B. pertussis* é o acúmulo de SNPs e posterior expansão clonal de linhagens mais adaptadas. A maior parte desses SNPs está presente em genes de virulência (BART et al., 2014; VAN GENT et al., 2012a).

*B. pertussis* apresenta inúmeros fatores de virulência envolvidos na infecção e progressão da doença, sendo as principais adesinas: filamentos de hemoaglutinina (FHA) (LOCHT et al., 1993), fimbrias (FIM) (WILLEMS et al., 1993), pertactina (PRN) (CHARLES et al., 1989) e fator de colonização da traquéia (tfc) (FINN; STEVENS, 1995) e as principais toxinas: toxina adenilato ciclase (ACT) (HEWLETT et al., 1976), citotóxina traqueal (TCT) e toxina pertussis (PTX), seu principal fator de virulência (LOCHT; KEITH, 1986). Estes fatores de virulência, assim como outros são expressos em diferentes momentos durante a infecção, e são regulados por um sistema chamado *bvg* (*Bordetella virulence gene*) (ARICÒ et al., 1991), composto de duas proteínas BvgA e BvgS. BvgS possui a capacidade se autofosforilar e fosforilar BvgA. BvgA quando fosforilada se liga ao DNA, promovendo ou inibindo a transcrição de determinados genes (MERKEL et al., 2003). Os genes de virulência estão dispersos ao longo do cromossomo da bactéria, porém alguns desses encontram-se em clusters

como é o caso dos genes do regulon *bvg*, *fha* e genes acessórios *fim* (Figura 4). Cinco desses fatores de virulência estão presentes em diferentes tipos de vacina acelulares, sendo elas PRN, FIM2, FIM3, PTX e FHA (MOOI; VAN DER MAAS; DE MELKER, 2013a).

Classes dos fatores de virulência

- Aderência
- Endotoxina
- Motilidade
- Protease
- Regulação
- Sistema de secreção
- Resistência ao soro
- Toxina

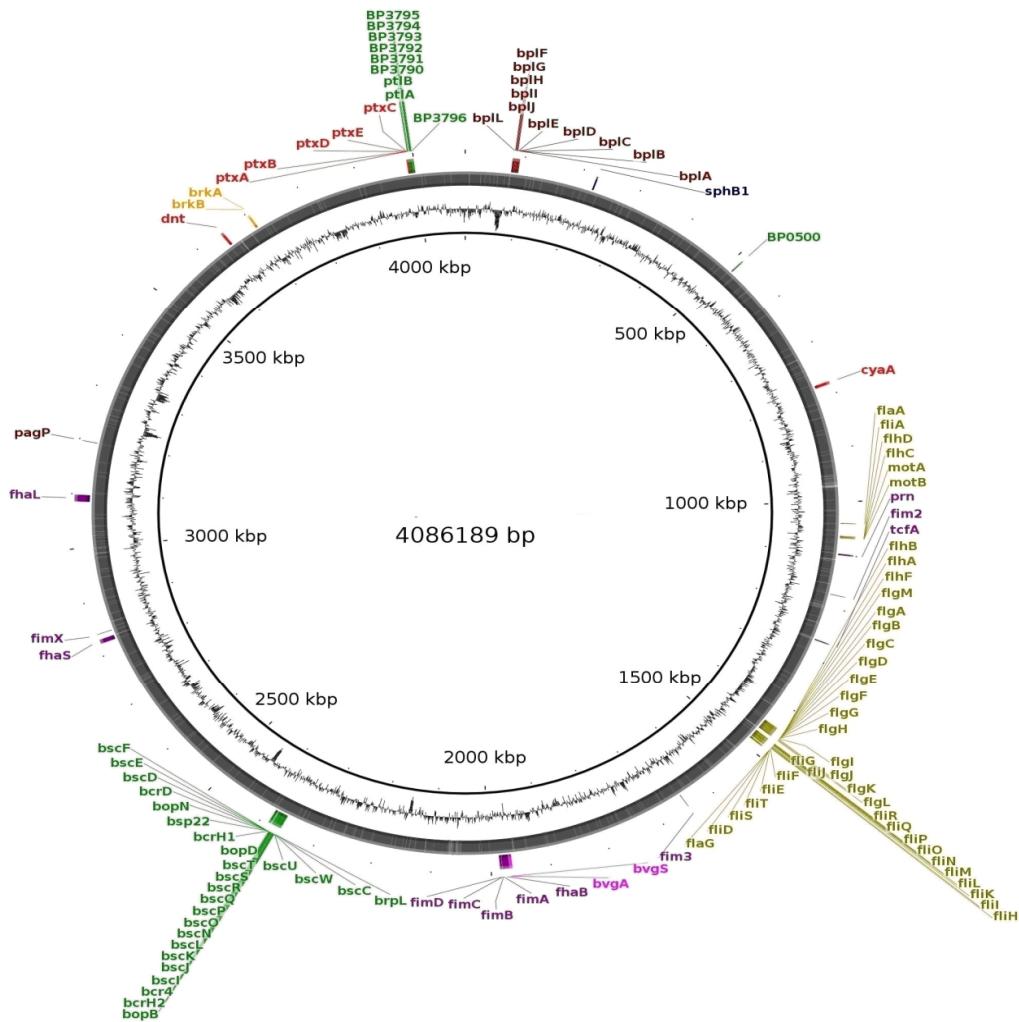


Figura 4. Distribuição de diferentes classes de genes ao longo do genoma de *B. pertussis*.

Estudos genéticos das populações de *B. pertussis* focam nos genes codificadores dessas proteínas, especialmente nos genes *ptxA* (que codifica a subunidade 1 da toxina pertussis), *prn* (que codifica a pertactina) e *fim3* (que codifica a fímbria do sorotipo 3), além da região promotora do *operon* da toxina pertússis (*ptxP*). Devido à presença de regiões repetitivas, o gene que codifica FHA é menos estudado. Vários trabalhos demonstram que o perfil alélico e a frequência de determinados perfis vêm variando a partir da introdução vacinação nos 50, caracterizando os períodos pré



e pós vacinais. (CASSIDAY et al., 2000; MOOI et al., 1998; OCTAVIA et al., 2012; POYNTEN et al., 2004; TSANG et al., 2004; VAN GENT et al., 2011).

As cinco subunidades que compõem a toxina PTX são codificadas por um conjunto de cinco genes (*ptxA-E* também chamados de *ptxS1-S5*), presentes em um *operon*. PtxA contém a atividade catalítica e tóxica sendo a mais imunogênica (DE MAGISTRIS et al., 1989) e polimórfica (VAN LOO; MOOI, 2002). Oito alelos já foram identificados para esse gene, quatro deles redundam numa mesma sequência de aminoácidos (MOOI, 2010). Para o promotor do operon *ptx* (*ptxP*) já foram descritos 18 alelos, sendo que o alelo *ptxP3* está associado a uma expressão elevada do gene *ptx* (DE GOUW et al., 2014; KING et al., 2013).

O gene que codifica a fímbria sorotipo 3 (*fim3*) possui cinco alelos descritos *fim3A, fim3B, fim3C, fim3A\** e *fim3E* também chamados de *fim3-1, fim3-2, fim3-3, fim3-4* e *fim3-5*, os quais codificam 4 proteínas não redundantes (MOOI, 2010).

Treze alelos para o gene *prn* foram identificados, e enquanto nos genes descritos anteriormente a variação mais observada são polimorfismos de base únicas (SNPs), no gene *prn* além de SNPs, também são observadas variações em duas regiões repetitivas que são imunodominantes e estão associadas a produção de anticorpos protetores em modelo murino (CHARLES et al., 1991; KING et al., 2001) (Figura 5). Recentemente isolados de *B. pertussis* que não expressão o gene *prn* foram identificados no Japão (OTSUKA et al., 2012), Estados Unidos (QUEENAN; CASSIDAY; EVANGELISTA, 2013) e Austrália (LAM et al., 2014). A não expressão desse gene está relacionada à sua interrupção por uma sequência de inserção, ou por um SNP que ocasiona em um códon de terminação e até a deleção do gene. Foi também demonstrado através de modelos murinos que as variantes PRN2 e PRN3 possuem uma maior adaptabilidade na colonização do trato respiratório quando comparada a PRN1 (VAN GENT et al., 2011).



**Figura 5. Alinhamento múltiplo mostrando o gene *prn*.** As caixas representam as regiões mais variáveis. Destacado em azul e amarelo estão as regiões repetitivas. Adaptado de Mooi et al 2013 (MOOI; VAN DER MAAS; DE MELKER, 2013b)

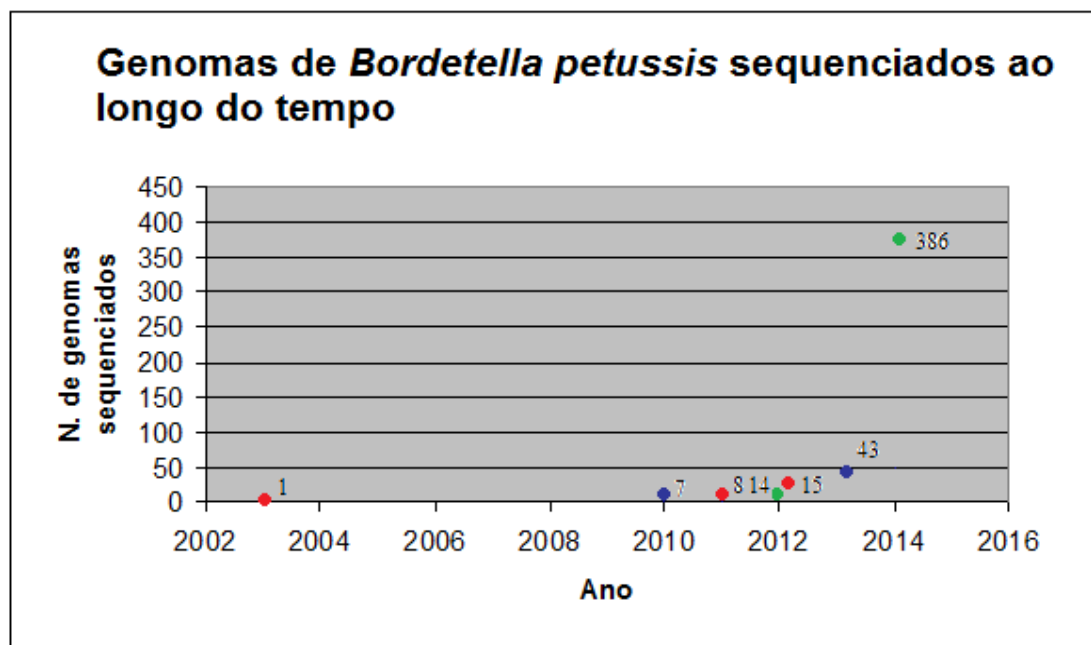
Inferências filogenéticas baseadas na análise de multilocus (MLST) utilizando-se desses genes apontam que isolados do período pré e pós vacinal formam clados distintos, e que isolados do período pré vacinal pertencem ao mesmo clado de cepas utilizadas para a produção de vacina. Isto sugere que essa mudança na frequência de alelos desses genes é consequência da introdução da vacina e da adaptação da bactéria frente à mesma (BART et al., 2014; VAN GENT et al., 2012b). Além de apresentar polimorfismos nesses genes, foi demonstrado que cepas do período pós vacinal que apresentam o alelo 3, para o promotor da toxina pertussis (*ptxP3*) apresentam uma expressão elevada do gene codificador para a toxina pertussis assim como de outros fatores de virulência (DE GOUW et al., 2014; KING et al., 2013).

Partindo de diferenças genotípicas pontuais, informações obtidas através do genoma completo permitiram que novas inferências a respeito da biologia e evolução de *B. pertussis* fossem feitas, assim como o surgimento de novas perguntas.

### 1.5. Histórico da genômica de *B. pertussis*

Em 2003, *Bordetella pertussis* teve seu genoma completamente sequenciado pela primeira vez por Parkhill e colaboradores (PARKHILL et al., 2003). Nesse mesmo estudo foram sequenciados os genomas das outras duas representantes das

*Bordetellas* clássicas (*B. bronchiseptica* e *B. parapertussis*). O sequenciamento revelou que *B. pertussis* possui um genoma de aproximadamente 4Mb (Mega bases) e conteúdo GC (guanina/citosina) de 67.72%. Desde o primeiro genoma sequenciado em 2003, o número de genomas de *B. pertussis* sequenciados vem aumentando consideravelmente: 6 novos *drafts*<sup>1</sup> foram publicados em 2010 por Bart e colaboradores (BART et al., 2010). Zhang e colaboradores liberaram o segundo genoma completo de *B. pertussis* em 2011 (ZHANG et al., 2011). Van Gente e colaboradores disponibilizaram outros 6 genomas em 2012 no formato de *reads*<sup>2</sup>. No mesmo ano Park e colaboradores disponibilizaram o terceiro genoma completo de *B. pertussis* (PARK et al., 2012a). Em 2013, Havill e colaboradores disponibilizam 28 novos *drafts*, e por fim, em um estudo de 2014, Bart e colaboradores disponibilizam no formato de *reads* 343 novos genomas (Figura 6)



**Figura 6 – Representação gráfica do número de genomas de *B. pertussis* sequenciados desde o primeiro em 2003.** De vermelho estão destacados os genomas completos, em azul *drafts* e em verde *reads*.

<sup>1</sup> *Draft*- Genoma quase completo. Apresentado na forma de múltiplos *contigs*

<sup>2</sup> *Reads*- Leituras provindas de um sequenciador

Cada um desses estudos aborda diferentes tópicos a respeito de *B. pertussis*, e levanta diferentes questões a respeito de sua evolução, e epidemiologia. Maiores informações sobre estes estudos são descritas a seguir .

O estudo de Parkhill e colaboradores, de 2003, sequenciou a cepa Tohama I. Isolada no Japão em 1954, a cepa Tohama I vem sendo usada por anos como referência para estudos com *B. pertussis* e é também usada para a produção de vacina em diversos países. Foi o primeiro isolado de *B. pertussis* a ter o seu genoma completamente sequenciado (PARKHILL et al., 2003). A metodologia aplicada foi o sequenciamento de Sanger. O genoma gerado possui 4.086.189 pb (pares de bases) dispostos em um único *contig*<sup>3</sup>.

Mais tarde em 2010, no primeiro estudo de genômica comparativa da espécie *B. pertussis* Bart e colaboradores sequenciaram 6 novos isolados de *B. pertussis*, todos provindos de surtos ocorridos na Holanda. O objetivo desse estudo era comparar linhagens de *B. pertussis* do período pré-vacinal e linhagens do período pós-vacinal. Análises comparativas entre esses isolados mais o genoma referência Tohama I estimou que *B. pertussis* apresenta uma densidade de SNPs de 1 SNP para cada 8.675 bases, o que coloca *B. pertussis* como um dos patógenos humanos mais monomórficos (BART et al., 2010).

Este mesmo estudo utilizou 471 SNPs para fazer uma inferência filogenética, demonstrando que isolados do período pós vacinal são distintos do período pré vacinal, assim como nas análises filogenéticas onde são considerados apenas genes de virulências anteriormente citados. Além de diferenças pontuais, o estudo também aponta grandes regiões de diferença entre isolados dos diferentes períodos que já tinham sido observadas por micro arranjo (HEIKKINEN et al., 2007a; MAHARJAN et al., 2008).

Todos esses genomas foram sequenciados na plataforma 454. O tamanho dos genomas variam de 3,81 Mb a 4,06 Mb, e se apresentam na forma de múltiplos *contigs* (de 397 a 1596).

O segundo genoma completo de *B. pertussis* a ser publicado trata-se de uma cepa vacinal chinesa, *B. pertussis* CS. A tecnologia empregada para o seu sequenciamento combinou duas abordagens: sequenciamento de Sanger e

---

<sup>3</sup> *Contig*- Sequências de *reads* sobrepostos resultantes em uma sequência consenso

pirosequenciamento utilizando-se a plataforma 454. O genoma gerado possui 4.124.236 pb , disposto em um único *contig*. (ZHANG et al., 2011)

No ano de 2012, sete novos genomas foram publicados. Seis deles por van Gent e colaboradores, no qual é demonstrado que *B. pertussis* circulante na Holanda nos períodos de 1949 a 2010, passou por pelo menos quatro grandes pressões seletivas seguidas de expansão clonal das linhagens selecionadas, essas seleções estão associadas a mutações nos genes *ptxA*, *prn*, *fim3* e *ptxP*. Todos esses genomas foram disponibilizados no formato de *reads* (VAN GENT et al., 2012b). O outro genoma publicado em 2012, pertence à cepa ATCC ou 18323. Isolada nos Estados Unidos em 1946, essa cepa também é usada para testes de controle de qualidade da vacina contra pertússis. Teve seu genoma publicado em 2012, que foi sequenciado pelo método de Sanger. O genoma gerado possui 4.043.846 pb dispostos em um único *contig* (PARK et al., 2012a).

Em 2013, Harvill e colaboradores publicaram 28 genomas de *B. pertussis* relativos a surtos ocorridos nos Estados Unidos nos anos de 2010 a 2012. Todos os genomas foram sequenciados na plataforma Illumina e seus tamanhos variam de 3,83 Mb a 4,15 Mb , e se apresentam na forma de múltiplos contigs (de 25 a 285) (HARVILL et al., 2013).

Finalmente, em abril de 2014, Bart e colaboradores traçaram a estrutura global da população de *B. pertussis*. Utilizando-se de 343 genomas de isolados de todos os continentes, foi possível inferir que uma recente linhagem de *B. pertussis* se espalhou por várias partes do mundo e que boa parte das mutações ocorridas nos genes codificadores de proteínas que são compostos vacinais ocorreram após a introdução da vacina. O estudo trás isolados de todos os continentes, porém não há nenhum isolado do Brasil. Todos os 343 genomas estão disponíveis no formato de *reads*.

## **1.6 Papel da IS481 em *B. pertussis***

Sequências de inserção possuem um papel chave na evolução de *B. pertussis*. Quando seu genoma foi sequenciado pela primeira vez, foi demonstrado que na cepa Tohama I apresentava 238 sequências de inserção, sendo a maioria delas do tipo IS481 (PARKHILL et al., 2003). Presente apenas em *B. pertussis*, IS481 é responsável por inúmeros rearranjos, comuns no genoma de *B. pertussis* (PRESTON; PARKHILL;

MASKELL, 2004; STIBITZ; YANG, 1999). Além disso, sequências de inserção podem se inserir em um determinado gene, inativando-o, sendo um dos mecanismos de formação de pseudogenes em *B. pertussis*, que apresenta 358 sequências desse tipo. Isso é observado quando comparamos os genomas de *B. pertussis* e *B. bronchiseptica*, onde é possível observar que genes presentes em *B. bronchiseptica* estão agora inativos devido à inserção de IS481. Dessa maneira, acredita-se que o processo de perda gênica, responsável pela especiação em *B. pertussis* teve grande influência de sequências de inserção (HEIKKINEN et al., 2007a; PRESTON; PARKHILL; MASKELL, 2004). Mais recentemente, foram detectadas *B. pertussis* que não expressam o gene da pertactina (LAM et al., 2014; OTSUKA et al., 2012; QUEENAN; CASSIDAY; EVANGELISTA, 2013), e o principal mecanismo de inativação desse gene é a inserção de IS481. Como anteriormente mencionado, a pertactina é um dos principais compostos da vacina acelular, dessa maneira, variantes de *B. pertussis* pertactina-negativas podem apresentar uma vantagem frente a um ambiente vacinado. A presença de IS481 nessas variantes demonstra que seu papel na evolução de *B. pertussis* ainda é atual.

## **1.7. Justificativa**

A re-emergência de uma infecção imunoprevenível levanta questões a respeito da adaptação e evolução do patógeno. O cenário atual aponta que *Bordetella pertussis*, principal agente etiológico de pertússis, apesar de ser um dos patógenos mais monomórficos a ser caracterizado, tem sido capaz de causar doenças em países com alta cobertura vacinal. Apesar de a re-emergência de pertússis também ser observada no Brasil, pouco se sabe sobre linhagens de *B. pertussis* causantes de pertússis no país.

Portanto, para uma completa avaliação e caracterização destas linhagem(s) é de extrema importância avaliar o conteúdo do genoma completo para um melhor entendimento da re-emergência de pertússis no país.

# Objetivos

## **2. Objetivos**

### **2.1. Objetivo geral**

Caracterização genômica de *B. pertussis* do Brasil isolados nas regiões Sul, Nordeste e Norte do país nos anos de 2008, 2013 e 2004 respectivamente.

### **2.2. Objetivos específicos**

- Estabelecer o perfil alélico dos genes por MLST de genes de virulência (*prn, fim3, ptxA, ptxP*);
- Seqüenciar e avaliar o genoma completo de isolados de *B. pertussis* representativos no tempo (2004 a 2013) e espaço (regiões Sul, Nordeste e Norte do Brasil);
- Realizar epidemiologia genômica de *B. pertussis* através da genômica comparativa entre os genomas gerados neste estudo e genomas disponíveis em dados públicos.



# **Metodologia**

### 3. Metodologia

Apesar de *B. pertussis* ser um organismo geneticamente pouco variável, mudanças na frequência de alelos nos genes que codificam proteínas usadas na maior parte das vacinas acelular foram observadas em diversos países. Sendo assim, a primeira abordagem usada neste estudo foi a caracterização genética de nossos isolados por MLST, utilizando esses genes (*prn, fim3, ptxA, ptxP*) como marcadores de linhagens de *B. pertussis*.

Os resultados da caracterização genética por MLST foram usados para a uma reconstrução filogenética na qual são observadas as relações entre isolados brasileiros e de outras partes do mundo, além de terem sido determinadas as linhagens circulantes no país. Linhagens representativas foram escolhidas com base nos resultados filogenéticos e sua localidade geográfica, e tiveram o seu genoma completamente sequenciado. A informação obtida no sequenciamento foi usada para estudos genômicos. Genes relacionados a resistência a macrolídeos também foram sequenciados e analisados em todas os isolados deste estudo.

Do ponto de vista genômico foram analisados genes ortólogos entre isolados brasileiros e genomas disponíveis no *GenBank*, esses genes foram usados para análises de filogenia. Foram determinados, também, genes únicos e sintonia entre os isolados brasileiros sequenciados.

Considerando o papel da sequência de inserção IS481 na história evolutiva de *B. pertussis* também foram determinadas as regiões do genoma que foram interrompidas por esta inserção e que estão íntegras no genoma referência (Figura 7).

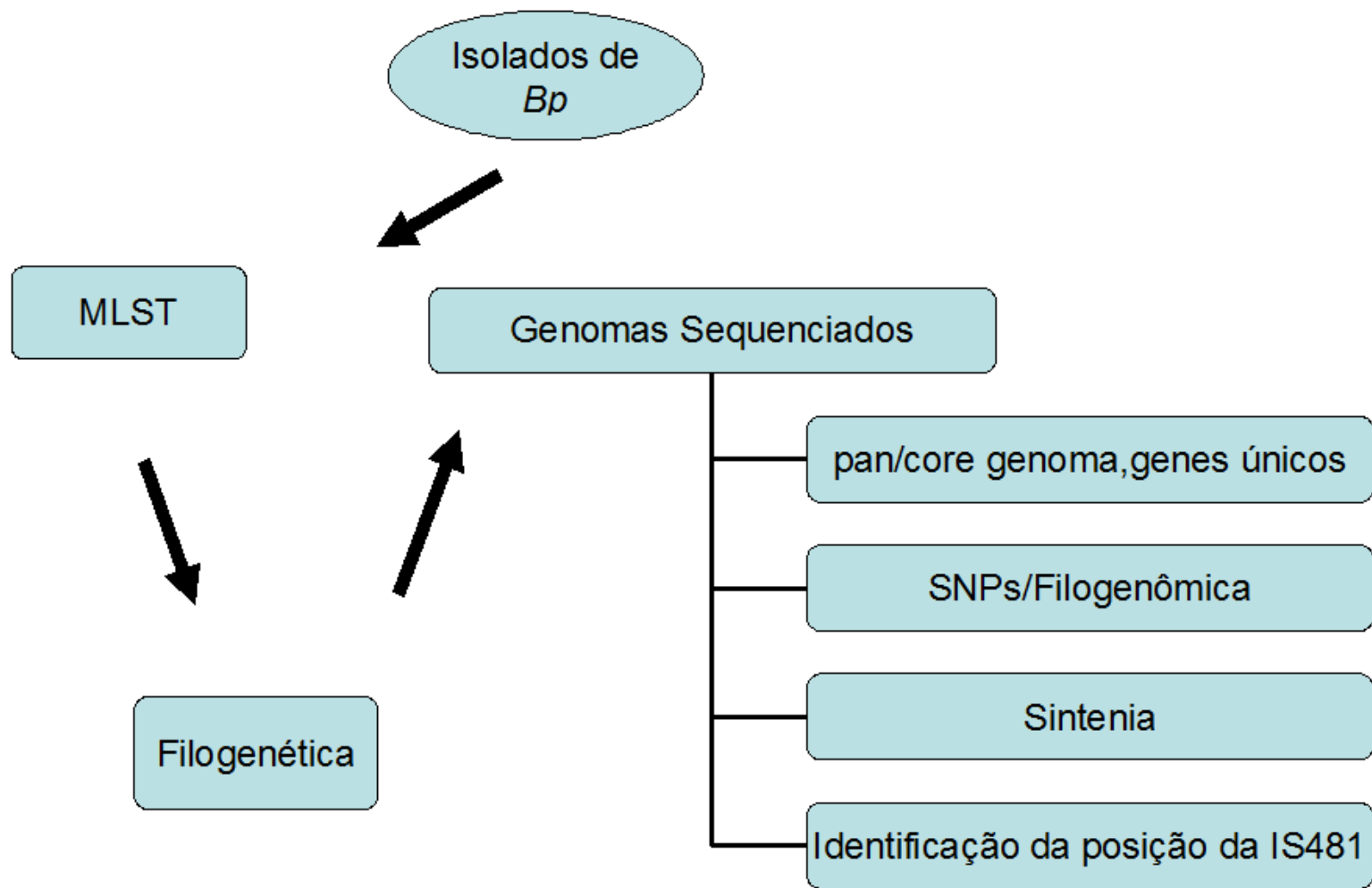


Figura 7. Fluxograma com a metodologia aplicada neste estudo.

### 3.1. Trabalho Experimental

#### 3.1.1. Isolados de *B. pertussis*

Fizeram parte dessas análises 26 isolados de *B. pertussis* pertencentes à Coleção de Bactérias do Meio Ambiente e Saúde (CBMA) do Laboratório de Genética Molecular e Microrganismos (LGMM), obtidos de casos clínicos de pertussis ocorridos nos estados de Alagoas (AL), Pernambuco (PE), Pará (PA) e Rio Grande do Sul (RS), entre os anos de 2004 e 2013 (Tabela 2).

O cultivo das bactérias foi feito primeiramente em meio MCS de acordo com Ohtsuka e colaboradores (2009). Posteriormente o inóculo foi repicado em ágar Bordet-Gengou, com acréscimo de sangue de cavalo, segundo as recomendações do fabricante HIMEDIA, ou ágar carvão, seguindo as instruções do fabricante OXOID, substituindo-se sangue por soro fetal bovino, e foram incubados a 35°C por pelo menos cinco dias.

**Tabela 2. Isolados De *B. pertussis* usados neste estudo, sua localidade e ano de isolamento.**

<b>Isolado</b>	<b>Estado</b>	<b>Ano</b>
MCSS	AL	2008
4SB,EBF,LBF,5SB,AFS ALSB,LJCMM,ESB	AL	2009
100,104,106,110,120,123 124,142,148,154,156,167	RS	2008
07	RS	2009
Bp168,Bp181	PE	2013
Bp2041,Bp2969	PA	2004

AL: Alagoas, RS: Rio Grande do Sul, PE: Pernambuco, PA: Pará

### 3.1.2. Extração de DNA, amplificação e sequenciamento dos genes para o MLST

Extração de DNA foi realizada pelo método de choque térmico onde massa bacteriana de cada amostra foi transferido para tubos do tipo “Eppendorf” adicionado de 500 µL de solução salina 0,85% e centrifugado a 3000 rpm por 5 minutos. Essa etapa foi repetida mais uma vez. O precipitado foi então ressuspensão em água mili-Q estéril e a suspensão submetida a banho-maria fervente (aproximadamente 100°C) por 10 minutos, com subsequente congelamento a -20°C. Esse material foi posteriormente descongelado à temperatura ambiente e centrifugado rapidamente (1400 rpm por 15 segundos) para ser utilizado nas reações de PCR.

Os iniciadores utilizados para amplificação dos genes *prn* (pertactina), *fim3* (fímbria sorotipo 3), *ptxS1* (subunidade 1 da toxina pertússis) e *ptxP* (promotor do operon da toxina pertússis) estão listados na Tabela 3. Para o gene 23s relacionado à resistência a macrolídeos a metodologia aplicada foi a mesma de Guillot e colaboradores (2012), e para os genes *gyrA*, *gyrB*, *parC*, e *parE* relacionados a resistência a quinolonas, a metodologia aplicada foi a mesma de Ohtsuka e colaboradores (2009). Para a amplificação de regiões específicas do genoma, a reação de PCR foi realizada em um volume final de 50 µL. As reações individuais foram compostas de água mili-Q estéril, tampão de reação 1X (Promega), 3 mM de MgCl<sub>2</sub> (Promega), 0,2 mM de dNTP (*ultrapure dNTP set* [dATP, dCTP, dGTP, dTTP]- Amersham), 200 ng de cada iniciador, 1,5 unidades da enzima Taq DNA Polimerase (Promega) e 5 µL do DNA total, obtido por choque térmico.

A reação foi realizada nas seguintes condições: pré-desnaturação a 94° por 5 minutos, seguida por 40 ciclos a 94°C por 30 segundos, 55°C por 30 segundos e 72°C com duração variando de acordo com o tamanho do alvo, considerando a relação 1 minuto/1 kb. A amplificação pela reação de PCR foi visualizada através de corrida de 10 µL do *amplicon* em gel de agarose a 1,5% em TAE 1X.

Os fragmentos amplificados foram purificados com QIAquick PCR Purification Kit – QIAGEN, e fragmentos extraídos do gel de agarose foram purificados com QIAquick Gel Extraction Kit – QIAGEN. Posteriormente, os produtos purificados foram utilizados como molde de DNA para a reação de sequenciamento de suas duas fitas com o sistema comercial BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems – Perkin Elmer Co) no sequenciador automático 3730 XL (Life Technology Applied) pertencente à Plataforma de Sequenciamento PDTIS (FIOCRUZ). Esta técnica já está padronizada e vem sendo utilizada rotineiramente pelo laboratório de Genética Molecular de Microrganismos, do IOC. As análises das sequências foram realizadas com auxílio dos pacotes de programas de análise de sequências Chromas (disponível em <http://goo.gl/aKSvgD>), Seqman (SWINDELL; PLASTERER, 1997), BioEdit (HALL., 1999) e MEGA5 (v.5.3) (TAMURA et al., 2011). As sequências foram comparadas àquelas depositadas no banco de sequências *GeneBank* (números de acesso disponíveis na tabela suplementar 1 de Bart e colaboradores (2014) ).

**Tabela 3- Iniciadores para amplificação e/ou sequenciamento de genes de virulência para caracterização genética.**

<b>Iniciador</b>	<b>Sequência</b>	<b>Referência</b>
PRNA-F	GCCAATGTACGGTCCA	(MOOI et al., 1998)
PRNA-R	GCAAGGTGATCGACAGGG	(MOOI et al., 1998)
PRNB-F	AGCTGGGCGGTTTCAGGT	(MOOI et al., 1998)
PRNB-R	CCGGATTCAGGCGCAACTCA	(MOOI et al., 1998)
PTXP-F	AATCGTCCTGCTCAACCGCC	(MOOI et al., 2009)
PTXP-R	GGTATACGGTGGGCGGGAGGA	(MOOI et al., 2009)
PTXS1-F	CCCCTGCCATGGTGTGATC	(FIETT et al., 2003)
PTXS1-R	TCAATTACCGGAGTTGGGCG	(FIETT et al., 2003)
FIM3-F	GACCTGATATTCTGATGCCG	(BORISOVA et al., 2007)
FIM3-R	AAGGCTTGCCGGTTTTTTTTGG	(BORISOVA et al., 2007)
PTXS1sq-F*	GTGGGGAAACAACGACAA	Este estudo
PTXS1sq-R*	GTGTAGGGGTTGGATTG	Este estudo
Fim3sq-F*	CACCCTCAACCATATCAA	Este estudo
Fim3sq-R*	TCTTGCTGCCATTGGTGA	Este estudo

\*Iniciadores usados apenas para sequenciamento

### **3.2. Análise de seqüências de multilocus (MLST – Multilocus Sequence Analysis)**

Com base na literatura (CASSIDAY et al., 2000; MOOI et al., 2009) foram selecionados os genes *prn* (pertactina), *fim3* (fímbria sorotipo 3), *ptxS1* (subunidade 1 da toxina pertussis) e *ptxP* (promotor do operon da toxina pertússis) para análise de MLST. O alinhamento das seqüências concatenadas foi realizado com o programa MAFFT v6.850 (KATOH; TOH, 2008).

As seqüências concatenadas foram usadas para a construção de uma árvore filogenética usando o método máxima verossimilhança do programa PhyML (GUINDON et al., 2010) com os seguintes parâmetros conforme Bart e

colaboradores (2010): modelo evolutivo HKY; transição/transversão: estimada; proporção de sítios não-variáveis: estimada; número de categorias de taxa de substituição relativa: 4, parâmetro de substituição gamma: estimado; árvore inicial: BIONJ; otimizar topologia da árvore: sim; otimizar comprimento dos ramos e parâmetros de taxa: sim.

### **3.3. Sequenciamento do genoma utilizando a plataforma GS 454 Junior**

Quatro dos 26 isolados foram selecionados para ter o seu genoma completamente sequenciado. A seleção foi feita com base nas análises de MLST e localização geográfica dos isolados. Os isolados Bp168, Bp100, Bp2969 e Bp1913 foram seqüenciados através da tecnologia do GS 454 Junior, que pertence a Plataforma de Sequenciamento de Alto Desempenho do Instituto Oswaldo Cruz. O DNA genômico foi extraído através do *kit ChargeSwitch® Mini Bactéria Kit* do fabricante invitrogen<sup>TM</sup>, e reação de sequenciamento foi feita conforme indicado pelo fabricante Roche.

Os genomas foram montados através do *software* de análises *Newbler* disponibilizado pela empresa Roche, e anotados mediante o servidor web RAST (*Rapid Annotation using Sybsystems Technology*) (AZIZ et al., 2008) Os seis genomas publicados no estudo de Bart e colaboradores (2010) também foram reanotados através do RAST.

Antes de se trabalhar com informações obtidas de um sequenciamento de alto desempenho é de suma importância a análise dos dados brutos gerados pela máquina, incluindo a aferição da qualidade das bases em cada *read* e cálculo do tamanho médio dos *reads*. Foi gerado um *boxplot* com a distribuição da qualidade ao longo dos *reads* através da plataforma Galaxy do servidor do Centro de Bioinformática da Holanda (NBIC) (BLANKENBERG et al., 2010). Também foi gerado um gráfico com a distribuição do tamanho de cada *read* utilizando-se o pacote R (R CORE TEAM, 2013).

### **3.4. Identificação de genes ortólogos, genes únicos e genoma core de *B. pertussis* e filogenia**



Para a identificação dos genes ortólogos foi utilizado o *pipeline* GET\_HOMOLOGUES (CONTRERAS-MOREIRA; VINUESA, 2013), que permite a busca de ortólogos através de diferentes algoritmos de clusterização. O algoritmo de escolha foi o OrthoMCL (*e-value* de  $1e-50$ , identidade  $\geq 50\%$ , cobertura  $\geq 80\%$ ) (LI; STOECKERT; ROOS, 2003). Nessa análises foram desconsiderados genes identificados como inparálogos.

Foram incluídos 40 genomas. 6 genomas de *B.pertussis* isolados na Holanda nos períodos de 1949 a 2000 publicados no estudo de Bart e colaboradores (BART et al., 2010). 28 genomas de *B. pertussis* de isolados dos Estados Unidos no período de 2010 a 2012, publicados em um *genome announcement* por Harvill e colaboradores (HARVILL et al., 2013), o genoma de referência de *B. pertussis* Tohama I publicado no estudo de Parkhill e colaboradores (PARKHILL et al., 2003) e o genoma de *B. bronchiseptica* cepa RB50 publicado no mesmo estudo, além dos 4 genomas de *B. pertussis* sequenciados neste estudo (Tabela 4).

Os genes pertencentes a um genoma em particular que não tiveram ortólogos identificados em nenhum dos demais genomas eram considerados candidatos a possíveis genes únicos.

Uma vez o OrthoMCL utiliza-se da sequência de proteínas (proteoma) para a identificação de ortólogos, é necessário que a sequência de nucleotídeos que quando traduzida originará aquela proteína esteja anotado como gene codificador da mesma. Sendo assim, o proteoma originado de um genoma pode ser diferente dependendo do anotador usado para esse genoma. Desse modo, muitos genes considerados únicos (ou seja, sua sequência de aminoácidos não possuía ortólogo em nenhum outro genoma comparado) a um determinado genoma podem ser falsos positivos, uma vez que esse gene considerado único pode estar presente em outro genoma no qual esse gene não foi identificado pelo anotador de escolha. Por esse motivo, nas comparações de genes únicos foi realizado uma busca de similaridade por Blast para a validação dos mesmos.

Os genes que possuíam ortólogos presentes em todos os isolados (genoma core) foram concatenados, alinhados pelo programa MAFFT v6.850 (KATOH; TOH, 2008) e usados para usados para uma reconstrução

filogenética, que foi feita através do software UGENE (OKONECHNIKOV; GOLOSOVA; FURSOV, 2012) utilizando o método de Neighbour-joining (SAITOU; NEI, 1987), Kimura 2-parâmetros (KIMURA; OTA, 1972).

Para análises de genoma *core* e pangenoma foram incluídos apenas 17 genomas representativos, na seguinte ordem: Tohama I, 6 genomas de isolados da Holanda ordenados por data de isolamento (do mais antigo para o mais recente), 6 genomas de isolados dos Estados Unidos ordenados por data de isolamento (do mais antigo para o mais recente) e os 4 genomas gerados neste estudo ordenados por data de isolamento (do mais antigo para o mais recente) (Tabela 4).

O conteúdo gênico dos 4 isolados sequenciados neste estudo foi comparado entre si, juntamente com o genoma de referência Tohama I. A fim de se evitar viés de anotação, essa comparação foi repetida com o genoma de Tohama I re-anotado pelo servidor web RAST, o mesmo usado para anotar os genomas gerados neste estudo.

Com o auxílio de *scripts* em R e dos *scripts* em *Perl* **compare\_clusters** e **parse\_pangenome\_matrix** disponíveis no pacote de *scripts* do GET\_HOMOLOGUES foram gerados gráficos de *Venn* mostrando o compartilhamento de genes entre os isolados brasileiros e o genoma referência Tohama I antes e depois da anotação.

**Tabela 4. Genomas usados para as análises deste estudo**

<b>Genoma</b>	<b>Origem</b>	<b>Ano de isolamento</b>	<b>Nº de acesso GenBank</b>
Bp100	Rio Grande do Sul	2008	Este estudo
Bp168	Pernambuco	2013	Este estudo
Bp2041	Pará	2004	Este estudo
Bp1913	Pará	2004	Este estudo
B0558	Holanda	1949	ADKR00000000
B1193	Holanda	1950	ADKS00000000
B1831	Holanda	1999	ADKT00000000
B1834	Holanda	1999	ADKU00000000
B1917	Holanda	2000	ADKV00000000
B1920	Holanda	2000	ADKW00000000
2250905	Estados Unidos	2010-2012	AXSU00000000
2356847	Estados Unidos	2010-2012	AXST00000000
2371640	Estados Unidos	2010-2012	AXSS00000000
STO1-SEAT-0006	Estados Unidos	2010-2012	AXSR00000000
STO1-SEAT-0007	Estados Unidos	2010-2012	AXSQ00000000
STO1-CHLA-0011	Estados Unidos	2010-2012	AXSP00000000
H897	Estados Unidos	2010-2012	AXSO00000000
H918	Estados Unidos	2010-2012	AXSN00000000
H921	Estados Unidos	2010-2012	AXSM00000000
H939	Estados Unidos	2010-2012	AXSL00000000
H973	Estados Unidos	2010-2012	AXSK00000000
STO1-SEAT-0004	Estados Unidos	2010-2012	AXSJ00000000
I002	Estados Unidos	2010-2012	AXSI00000000
I036	Estados Unidos	2010-2012	AXSH00000000
I176	Estados Unidos	2010-2012	AXSG00000000
STO1-CHLA-0006	Estados Unidos	2010-2012	AXSF00000000
CHLA-15	Estados Unidos	2010-2012	AXSD00000000
CHLA-13	Estados Unidos	2010-2012	AXSE00000000
CHLA-20	Estados Unidos	2010-2012	AXSC00000000
CHLA-26	Estados Unidos	2010-2012	AXSB00000000
STO1-CHOC-0016	Estados Unidos	2010-2012	AXSA00000000
STO1-CHOC-0008	Estados Unidos	2010-2012	AXRV00000000
STO1-CHOC-0017	Estados Unidos	2010-2012	AXRZ00000000
STO1-CHOC-0018	Estados Unidos	2010-2012	AXRY00000000
STO1-CHOC-0019	Estados Unidos	2010-2012	AXRX00000000
STO1-CHOC-0021	Estados Unidos	2010-2012	AXRW00000000
STO1-CHOM-0012	Estados Unidos	2010-2012	AXRU00000000
STO1-CNMC-0004	Estados Unidos	2010-2012	AXSV00000000
Tohama I	Japão	1952	NC_002929.2
<i>B. bronchiseptica</i> RB50	--	--	NC_002927.3

Em azul genomas gerados neste estudo. Em vermelho genomas usados para análise de genoma *core* e pangenoma

### 3.5. Identificação de SNPs

Com o objetivo de uma melhor resolução acerca da filogenia de *B. pertussis*, considerando os isolados do Brasil, foi construída uma análise a partir dos SNPs identificados nos genomas.

A identificação de SNPs foi feita na plataforma *web* SNPsFinder (SONG et al., 2005), construída especificamente para genomas de microrganismos extremamente relacionados. SNPsFinder usa um genoma fornecido pelo usuário como referência, e mapeia este genoma contra os outros genomas fornecidos utilizando o algoritmo MegaBlast (ZHANG et al., 2000). Os resultados do MegaBlast são “parseados” e SNPs são dados em formato de tabela. Para este estudo foi usado o genoma referência de *B. bronchiseptica* RB50, e porcentagem de identidade para o busca de similaridade do MegaBlast foi de 95%. SNPs em genes ou regiões intergênicas presentes em todos os genomas, foram usados para a construção de uma árvore filogenética. Aqueles SNPs exclusivos de *B. bronchiseptica* foram retirados da análise.

Os sítios correspondentes a 826 SNPs encontrados foram concatenados e usados para a construção de um árvore de máxima verossimilhança pelo programa PhyML, com os seguintes parâmetros, de acordo com Bart e colaboradores (2010): modelo evolutivo HKY; transição/transversão: estimada; proporção de sítios não-variáveis: estimada; número de categorias de taxa de substituição relativa:4,parâmetro de substituição gamma: estimado; árvore inicial: BIONJ; otimizar topologia da árvore: sim; otimizar comprimento dos ramos e parâmetros de taxa: sim (BART et al., 2010).

### 3.6. Sintenia

Apesar de ser os genomas de *B. pertussis* serem conservados considerando às sequências nucleotídicas, estes são bastante dinâmicos quando se analisa a ordem gênica (HARVILL et al., 2013; STIBITZ; YANG, 1999). Rearranjos são comuns mesmo em linhagens próximas. Estes

rearranjos são frequentemente flanqueados pela sequência de inserção IS481 que está presente em numerosas cópias ao longo de todo o genoma.

A avaliação da sintonia entre os genomas dos isolados brasileiros foi feita em duas etapas. Uma vez que os genomas gerados estavam na forma de múltiplos *contigs*, os mesmos foram reordenados através da aplicação Mauve (RISSMAN et al., 2009). A referência escolhida foi o genoma de *B. bronchiseptica* RB50. Após reordenados, os 4 genomas foram mapeados entre si, dois a dois, com uso do BLAST (ALTSCHUL et al., 1990), nas seguintes combinações Bp100 *versus* Bp168, Bp168 *versus* Bp2041, Bp2041 *versus* Bp1913. A visualização da sintonia então foi feita na aplicação ACT (Artemis Comparison Tool) (CARVER et al., 2005).

### **3.7. Identificação de regiões do genoma interrompidos pela IS481**

Considerando o papel da IS481 na história evolutiva de *B. pertussis*, foi também feita uma análise na tentativa de se descobrir quais genes ou regiões dos genomas dos isolados do Brasil estariam interrompidos por essa sequência de inserção em relação ao genoma referência Tohama I.

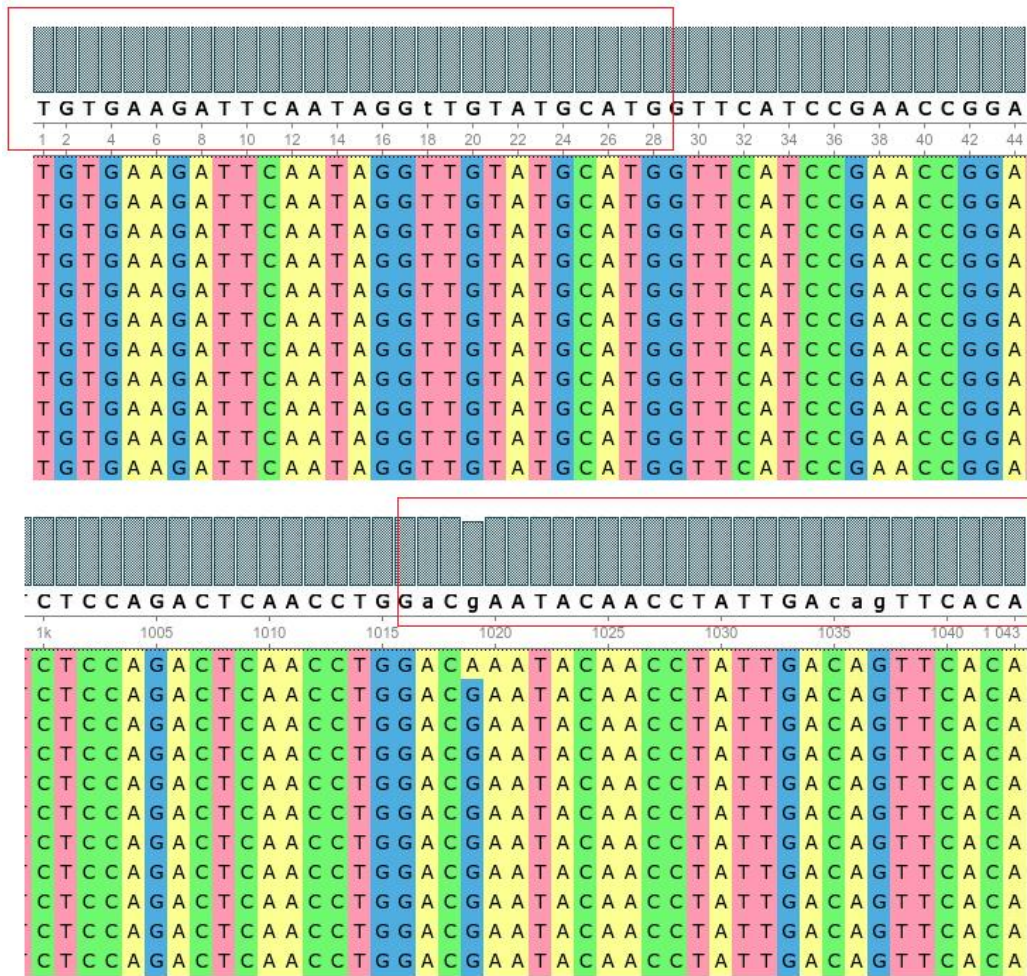
A sequência de inserção IS481 é uma sequência repetitiva que possui um tamanho de aproximadamente 1047 pares de bases. Considerando que *reads* de genomas sequenciados através da plataforma GS 454 Junior não ultrapassam em média os 500 pares de bases, a identificação desta sequência de inserção em *B. pertussis* em genomas provenientes desse tipo de sequenciamento se torna um desafio. Foi desenvolvida, então, uma abordagem simples que tenta solucionar parcialmente esta limitação.

A identificação de regiões interrompidas pela sequência de inserção IS481 consistiu nas seguintes etapas:

Primeiramente todas as sequências de inserção presentes em Tohama I foram recuperadas e posteriormente alinhadas pelo programa MAFFT disponível na aplicação UGENE. O visualizador de alinhamento do UGENE demonstra graficamente posições conservadas em todo o alinhamento. Sendo assim, usamos os 28 nucleotídeos iniciais e finais de as IS481 como “isca” para a identificação de *reads* do sequenciamento, através de uma busca literal.

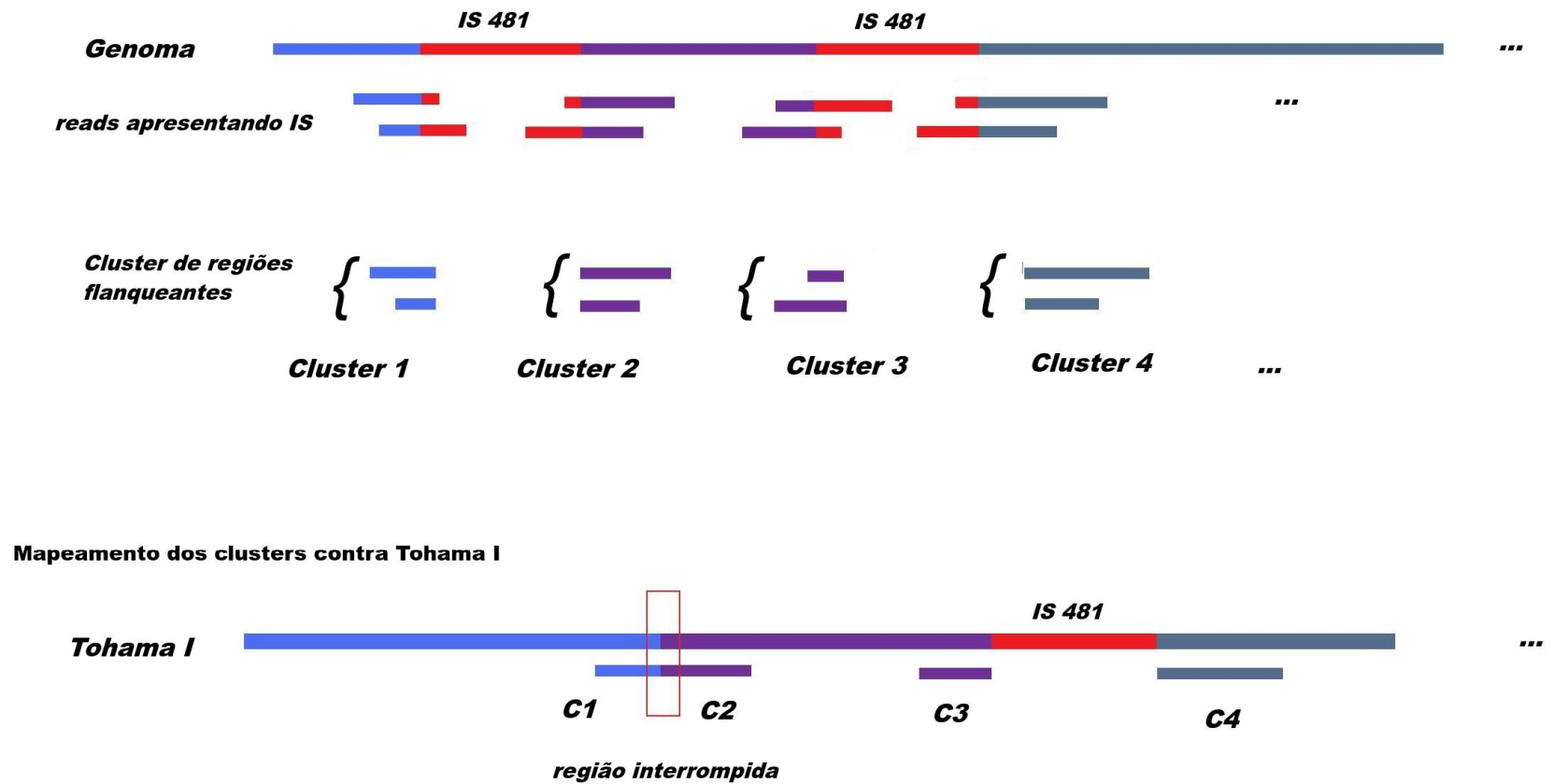
Quando uma posição não era 100% conservada em todas as IS481, foram utilizado um caractere coringa (Figura 8).

Foram criados dois grupos de *reads*: um grupo com *reads* que continham a IS481 e a região 3´ adjacente e um outro grupo de *reads* com a IS481 e a região 5´ adjacente. Posteriormente a região correspondente a IS481 foi deletada desses *reads*, sobrando apenas a região que estava adjacente a IS. Uma vez que muitas dessas sequências pertenciam a uma mesma região do genoma, foi usado o programa *cd-hit-est*, disponível no pacote de programas do CD-HIT (FU et al., 2012), para a clusterização dessas sequências. O percentual de identidade escolhido foi de 90%.



**Figura 8. Representação parcial do alinhamento de todas as sequências de Inserção IS481 de Tohama I.** Abaixo da barra cinza é mostrada a sequência consenso onde em maiúsculo estão regiões 100% conservadas. A caixa vermelha indica os nucleotídeos usados como “isca” para se buscar *reads* que continham essa sequência.

Os *clusters* encontrados foram então mapeados através do programa BWA-SW (LI; DURBIN, 2010) contra o genoma de *B. pertussis* Tohama I. A visualização desse mapeamento foi feita na aplicação UGENE, no qual foi possível observar quais regiões do genoma dos isolados brasileiros estavam separadas pela IS481, porém estão adjacentes no genoma de Tohama I (Figura 9).



**Figura 9. Representação da busca por regiões interrompidas pela IS481.** A região do genoma correspondente a IS481 está representada em vermelho. Regiões flanqueantes de *reads* que possuíam a “isca” usada foram clusterizadas e então mapeadas a Tohama I. Clusters que ficaram adjacentes representam regiões do genoma interrompidas em relação ao genoma de Tohama I.



Antes do início dessa abordagem, foi gerado um sequenciamento simulado na plataforma GS 454 Junior, com cobertura de 20x, do genoma *B. pertussis* Tohama I (semelhantes ao sequenciamento dos genomas usados este estudo). Este sequenciamento foi gerado através do programa ART (HUANG et al., 2012) e teve como objetivo testar a metodologia desenvolvida. Foram feitos os mesmos passos descritos anteriormente com os *reads* do sequenciamento simulado. Em seguida, *clusters* obtidos foram mapeados contra o genoma de *B. bronchiseptica* no qual foi possível observar regiões que estão contíguas no genoma de *B. bronchiseptica*, mas estão interrompidas em de *B. pertussis*, como observado anteriormente (PARKHILL et al., 2003).

A localização da região onde a IS481 está presente no genoma de isolados brasileiros foi comparada com dos outros genomas depositados no *GenBank*.

# Resultados

## 4. Resultados

### 4.1. Caracterização genética de isolados de *B. pertussis* do Brasil

#### 4.1.1. MLST de genes de virulência, e genes de resistência

A caracterização de isolados de uma espécie de bactéria aplicando a abordagem de MLST considera, geralmente, a combinação de genes, do genoma estável que são associados ao metabolismo da bactéria. Na espécie *B. pertussis*, mesmos entre isolados distintos no tempo-espaço, a variação nestes genes não é suficiente para estabelecer relações epidemiológicas entre eles. Nesta espécie, tem sido aplicado o MLST baseado nos genes de virulência (VAN LOO et al., 2002), e assim, foi demonstrado que *B. pertussis* de várias partes do globo vem variando desde a introdução das diferentes vacinas para pertússis a partir de 1950.

Neste estudo realizamos a caracterização genética de uma coleção de 26 *B. pertussis*, isoladas no Brasil entre 2004-2013, aplicando MLST a partir de 4 genes: pertactina, fimbria do sorogrupo 3, toxina pertussis e promotor do gene da toxina pertussis, *prn2-fim3-ptxS1-ptxP*, respectivamente.

A maioria dos isolados caracterizados (24 isolados) apresentou o perfil alélico *prn2-fim3B-ptxA1-ptxP3*, e foi detectado em todos os estados. Dois isolados apresentaram o perfil *prn2-fim3A-ptxA1-ptxP3*, distintos por um único SNP A/C no gene *fim3*. Este perfil só foi detectado 2 isolados do Pará (ano de 2004) (Figura 10, Tabela 5). Nenhuma mutação associada à resistência à quinolonas e macrolídeos foi encontrada.

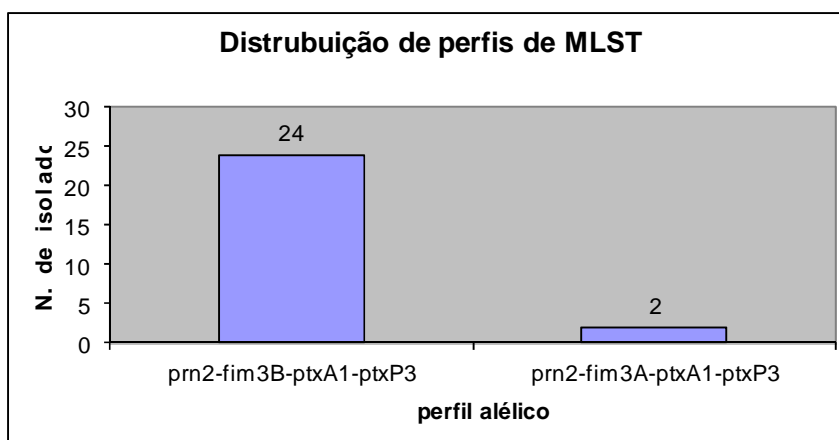


Figura 10. Distribuição dos perfis de MLST entre o conjunto de isolados de *B. pertussis*.

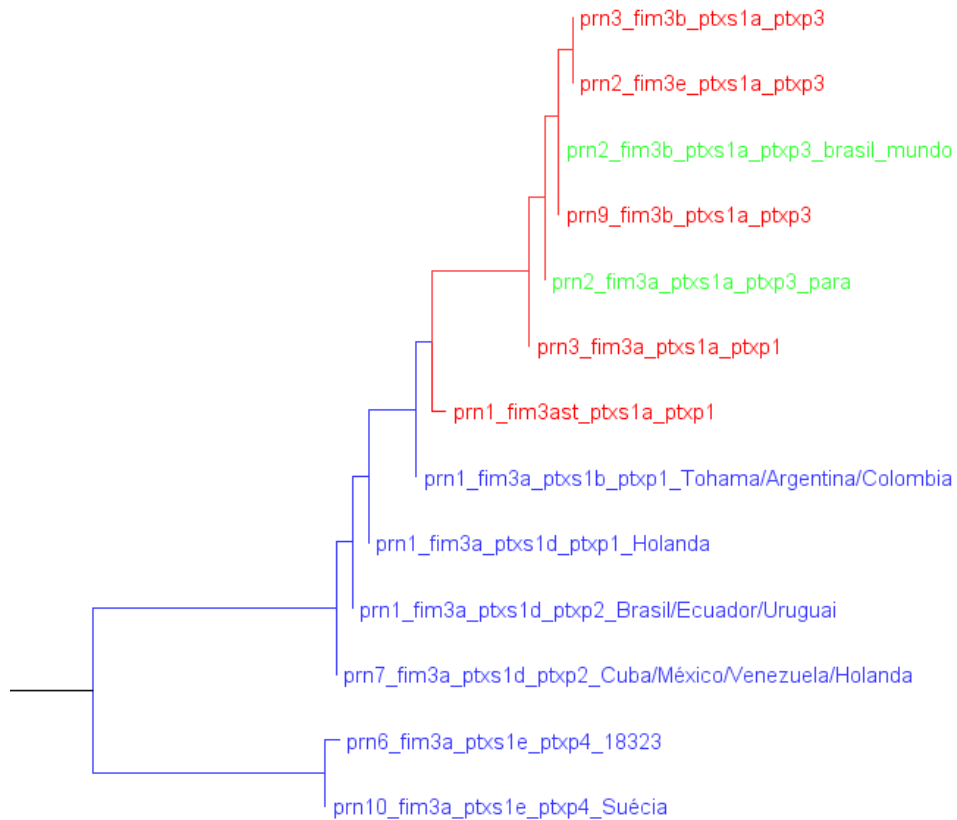
**Tabela 5. Isolados de *B. pertussis*, sua origem e perfil alélico dos genes de virulência.**  
Em verde estão destacados isolados com MLST distintos aos demais.

Isolado	Estado	Ano	GENES/ALELOS			
			<i>prn</i>	<i>fim3B</i>	<i>ptxS1</i>	<i>ptxP</i>
77	RS	2008	2	B	A	3
99	RS	2008	2	B	A	3
100 <sup>1</sup>	RS	2008	2	B	A	3
104	RS	2008	2	B	A	3
106	RS	2008	2	B	A	3
110	RS	2008	2	B	A	3
120	RS	2008	2	B	A	3
123	RS	2008	2	B	A	3
124	RS	2008	2	B	A	3
142	RS	2008	2	B	A	3
148	RS	2008	2	B	A	3
154	RS	2008	2	B	A	3
156	RS	2008	2	B	A	3
167	RS	2008	2	B	A	3
07	RS	2009	2	B	A	3
MCSS	AL	2008	2	B	A	3
EBF	AL	2009	2	B	A	3
LBF	AL	2009	2	B	A	3
5SB	AL	2009	2	B	A	3
ESB	AL	2009	2	B	A	3
AFS	AL	2009	2	B	A	3
ALSB	AL	2009	2	B	A	3
LJCM	AL	2009	2	B	A	3
4SB	AL	2009	2	B	A	3
Bp168	PE	2013	2	B	A	3
Bp181	PE	2013	2	B	A	3
Bp1913	PA	2004	2	B	A	3
Bp2041	PA	2004	2	A	A	3
Bp2969	PA	2004	2	A	A	3

RS: Rio Grande do Sul, AL: Alagoas, PE: Pernambuco, PA: Pará

#### 4.1.2. Relação genética entre isolados de *B. pertussis*

A relação genética entre isolados de *B. pertussis* de várias regiões do mundo, tanto dos períodos pré e pós-vacinal, foi determinada a partir da análise das sequências das quatro regiões, *prn-fim3-ptxS1-ptxP*, num total de 4390pb (Figura 11).



**Figura 11. Relações genéticas entre *B.pertussis* a partir das sequencias dos genes *prn-fim3-ptxS1-ptxP* concatenados (4390pb) usando o método de Máxima Verossimilhança. Destacadas em vermelho as linhagens pós-vacinais e em azul as linhagens vacinais. De verde encontram-se o perfil dos isolados do Brasil**

## 4.2. Genomas de *B. pertussis* do Brasil

Até o momento desta análise havia 37 genomas de *B. pertussis* disponíveis no GenBank, no qual 34 deles encontravam-se no formato de *drafts*. 28 desses genomas correspondiam a isolados de surto recente nos Estados Unidos (HARVILL et al., 2013), 6 correspondiam a isolados da Holanda dos períodos pré e pós vacinais (BART et al., 2010). Dos 3 genomas restantes, apresentados como genoma completo, este estudo utilizou apenas a cepa referência Tohama I (PARKHILL et al., 2003).

A análise por MLST revelou dois perfis distintos entre os isolados deste estudo: um perfil dominante, espalhado do norte ao sul do país, *prn2-fim3B-ptxS1A-ptxP3*, e um perfil somente encontrado em isolados de 2004 do estado do Pará, *prn2-fim3A-ptxS1A-ptxP3*. Considerando tempo/espaço e perfis de MLST, selecionamos o isolado Bp100, do Rio Grande do Sul (2008), o Bp168, de Pernambuco (2013) e Bp1913, do Pará (2004) todos pertencentes ao perfil dominante e o Bp2041, do Pará (2004), pertencente ao perfil minoritário, para terem seus genomas completos sequenciados e analisados, e desta forma revelarmos os primeiros genomas de *B. pertussis* do Brasil e América do Sul.

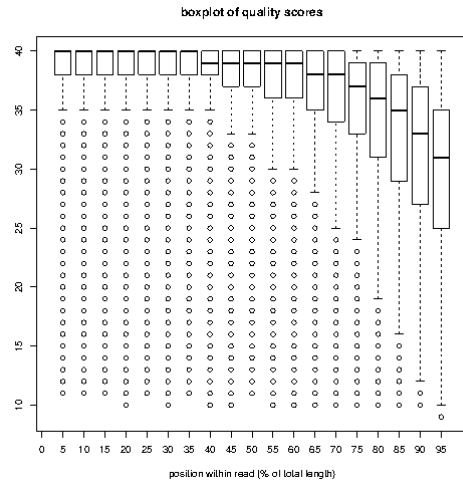
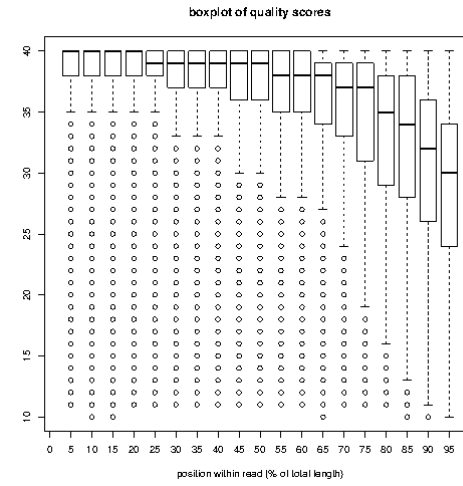
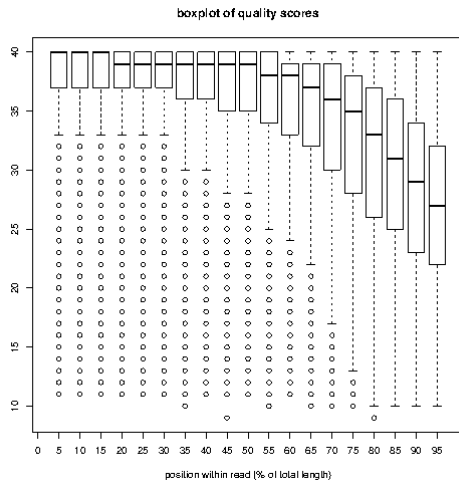
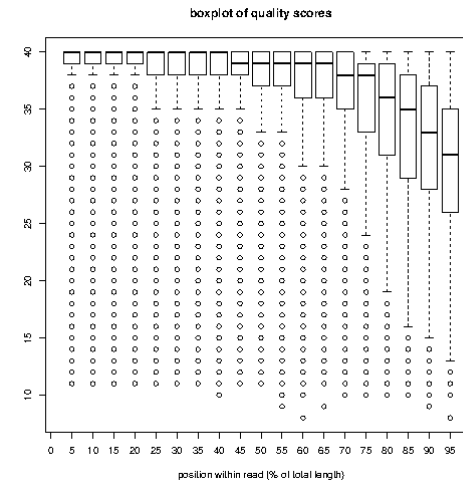
### 4.2.1. Métricas do sequenciamento genômico por alto desempenho

Dados brutos, obtidos pelo sequenciamento, realizado no sequenciador GS 454 Junior, foram analisados em relação ao número de bases sequenciadas, número de *reads*, tamanho médio dos *reads* e qualidade média do sequenciamento (Tabela 6, Figura 12-13).

**Tabela 6. Métrica do sequenciamento e montagem dos quatro genomas de isolados de *B. pertussis* do Brasil.**

Isolado	Número de <i>reads</i>	Tamanho médio dos <i>reads</i>	Número de bases sequenciadas	Qualidade média do sequenciamento	No de <i>contigs</i>	N50	Cobertura
Bp100	207466	453,793	94146691	36	277	21042	23X
Bp168	198984	458,41	91216212	35,5	275	20264	22X
Bp1913	178115	454,59	80969297	34,63	277	21027	20X
Bp2041	221443	464,671	102898185	36,29	280	20202	25X

O número de bases sequenciadas correspondeu ao esperado para uma corrida realizada em um GS-454-Junior (LOMAN et al., 2012), o que leva a uma cobertura maior que 20 vezes em todos os casos. A qualidade média do sequenciamento ficou próxima de 36 para os quatro genomas gerados, o que significa que há a probabilidade de 1 erro a cada 9000 bases seqüenciadas.

**A****B****C****D**

**Figura 12. Boxplot com a qualidade dos reads em relação a posição da base ao longo do read. É possível observar a queda de qualidade ao final de cada read, uma característica típica do sequenciamento por 454. – A: Bp168. B: Bp100. C: Bp1913. D: BP2041.**



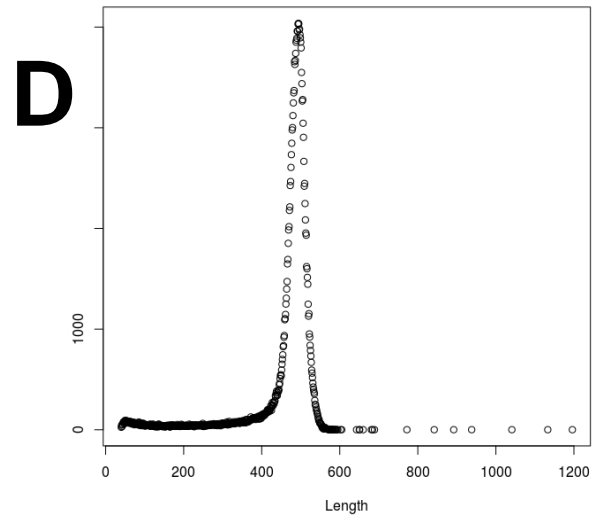
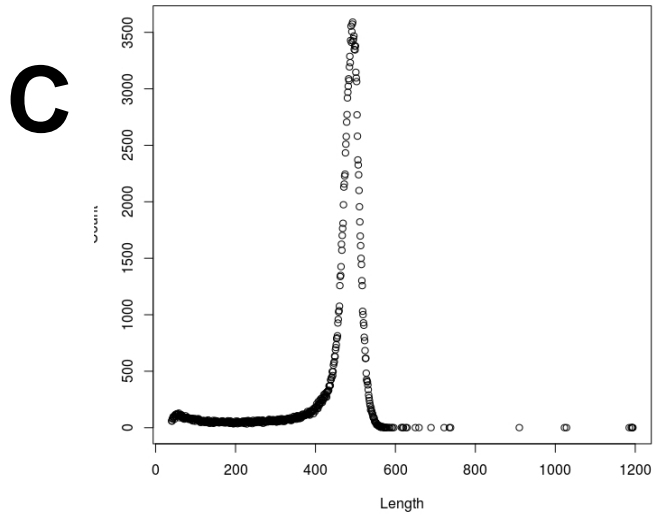
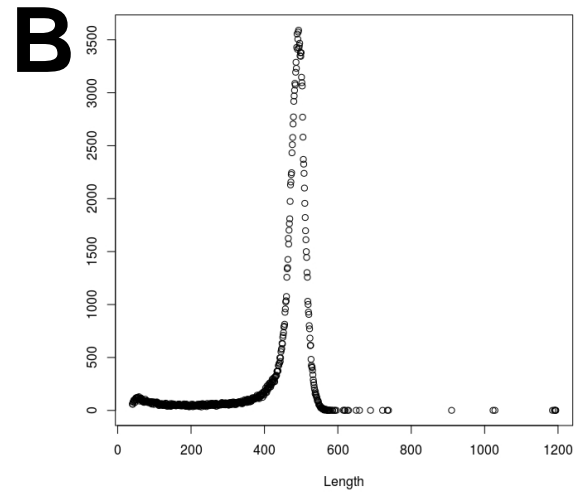
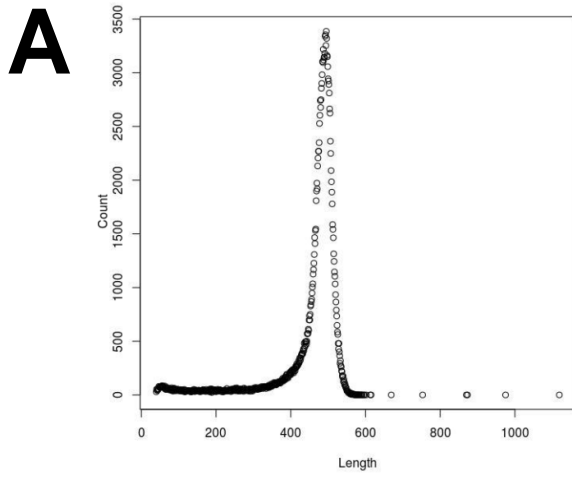


Figura 13. Gráfico com contagem do tamanho dos *reads*. A: Bp168, B: Bp100, C: Bp1913, D: B2041

#### 4.2.2. Características dos genomas sequenciados e anotados

Dos 40 genomas usados nas análises demonstradas a seguir, 11 deles foram re-anotados pelo RAST, sendo os 4 genomas gerados neste estudo, os 6 genomas do estudo de Bart e colaboradores (2010) e o genoma de referência Tohama I (PARKHILL et al., 2003). A descrição das características detalhadas dos genomas sequenciados por este estudo e anotados pelo RAST encontram-se na Tabela 7, e em detalhes no Anexo I.

**Tabela 7. Características dos genomas sequenciados neste estudo**

<b>Genoma</b>	<b>Tamanho</b>	<b>N. de <i>contigs</i></b>	<b>N. de sequências codificadoras</b>
Bp168	3.802.120 pb	275	3675
Bp100	3.818.795 pb	277	3702
Bp1913	3.822.732 pb	277	3771
Bp2041	3.804.142 pb	280	3628

As características dos genomas re-anotados podem ser encontradas no servidor do RAST, no endereço <http://rast.nmpdr.org/> através do *login*: bordpbrasil. senha: bordetellabrazil.

#### **4.2.3. Identificação de genes ortólogos, genoma core e pangenoma de *B. pertussis***

A partir de 28 *drafts* de isolados dos Estados Unidos obtidos de 2010 a 2012, 6 *drafts* de isolados da Holanda do período de 1949 a 2000, o genoma referência Tohama I, além dos 4 genomas dos isolados brasileiros e *B. bronchiseptica*, e aplicando o algoritmo OrthoMCL foram identificados 4448 *clusters/genes*. O genoma core destes 40 genomas é representado por 2050 genes (usados para a construção de uma árvore filogenética) (Figura 16). Após validação, não foi encontrado nenhum gene único em qualquer um dos isolados brasileiros.

Em relação ao genoma *core* e ao pangenoma, que foi estabelecido usando o genoma referência Tohama I, os 6 genomas de isolados da Holanda, 6 genomas representativos de isolados dos Estados Unidos (destacados na Tabela 4), e os 4 isolados do Brasil, é possível observar que em relação ao genoma *core* a curva segue a tendência de queda ao longo do gráfico, já em relação ao pangenoma observa-se duas grandes acentuações: uma após a adição do segundo genoma e outra após a adição do oitavo genoma (Figura 14).

# Pangenoma e Genoma core de *B. pertussis*

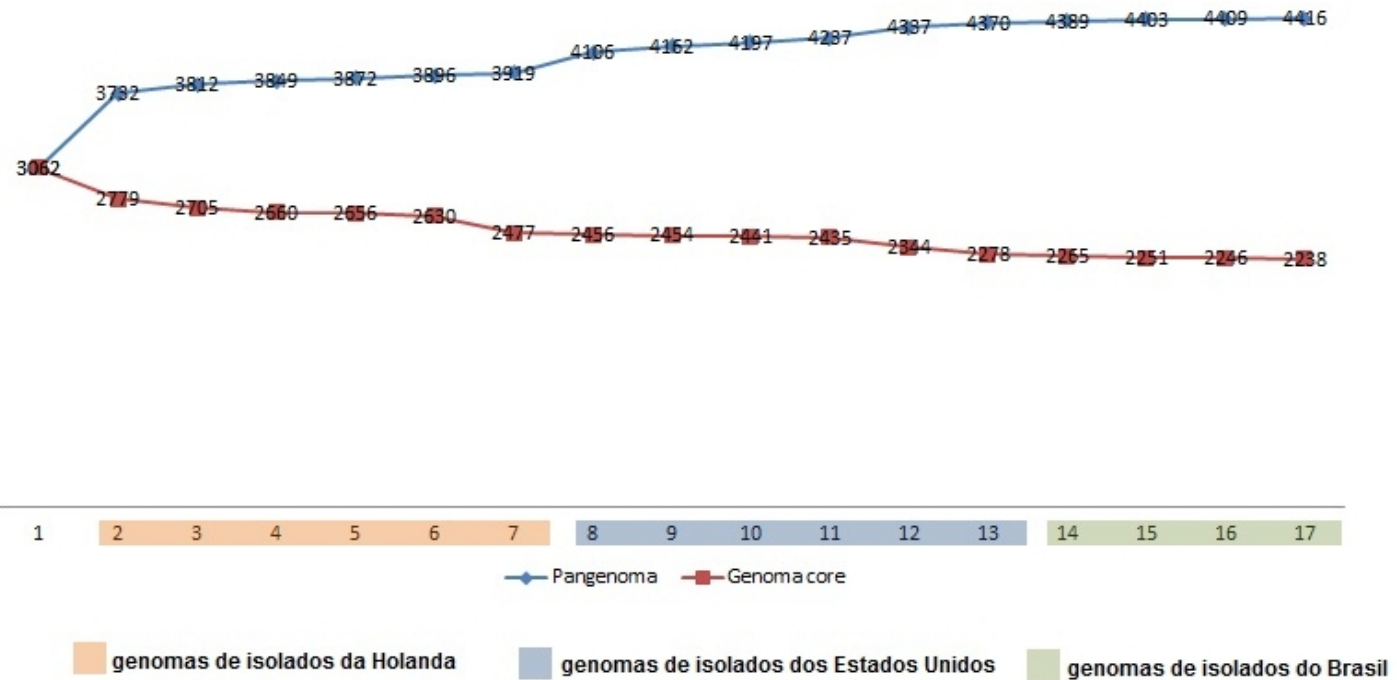
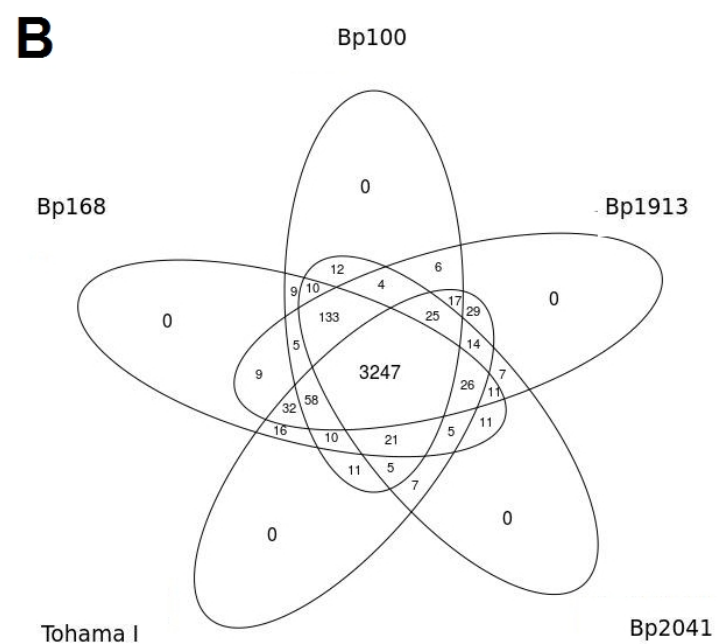
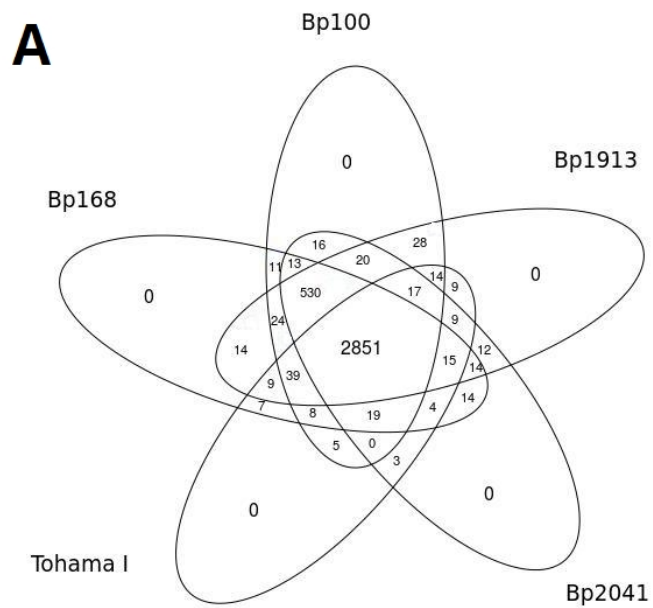


Figura 14- Representação do genoma core e pangenoma de *B. pertussis* a partir 17 genomas representativos no tempo-espaço

Foi realizada uma comparação entre o conteúdo gênico dos 4 isolados brasileiros e o da linhagem referência Tohama I. Foram gerados dois diagramas de Venn nos quais observamos o compartilhamento de genes entre estes 5 isolados, um com a anotação original de Tohama I, disponível no *GenBank* e outro com a re-anotação pelo RAST (este trabalho) o mesmo anotador usados para isolados brasileiros. No primeiro diagrama observa-se que Tohama I não possui nenhum gene único quando comparado aos isolados brasileiros, porém os isolados brasileiros apresentam 530 genes exclusivos. O genoma core desse conjunto é representado por 2851 genes (Figura 15 A). Já no segundo diagrama, usando o conteúdo da Tohama re-anotado pelo RAST, é possível notar que o número de genes únicos entre isolados brasileiros cai para 133 e o genoma core sobe para 3247 (Figura 15 B). Foi feita uma busca de similaridade por Blastn com estes 133 genes contra Tohama I, nos parâmetros padrão, para sua validação como genes únicos, dos quais 43 retornaram sem *hit*, e portanto únicos do conjunto do Brasil. A escolha do Blastn ao invés do blastx se deve ao fato de que dado a conservação do genoma de *B. pertussis*, dificilmente uma sequência protéica seria codificada por diferentes sequências de nucleotídeos. Estes genes estão discriminados na Tabela 8. A maioria destes genes codifica proteínas hipotéticas, porém foram encontrados genes relacionados a regulação gênica, transcrição, tradução, proteínas de membrana, entre outros.

**Tabela 8. Genes únicos dos genomas de isolados do Brasil em relação a Tohama I**

hypothetical protein
Branched-chain amino acid transport ATP-binding protein LivF
Branched-chain amino acid transport ATP-binding protein LivG
Branched-chain amino acid transport ATP-binding protein LivG
probable permease component of branched-chain amino acid transporter
probable permease component of branched-chain amino acid transporter
Putative translation initiation inhibitor, yjgF family
Acyl-CoA synthetases (AMP-forming)/AMP-acid ligases
putative branched-chain amino acid-binding protein
Transcriptional regulator, lclR family
probable short chain dehydrogenase
probable ATP-binding component of ABC transporter
putative inner membrane component of binding-protein-dependent transportsystem
hypothetical protein
ABC transporter substrate binding protein
Enoyl-CoA hydratase
L-carnitine dehydratase/bile acid-inducible protein
putative amidase
putative amidase
hypothetical protein
putative exportedprotein
putative membrane protein
hypothetical protein
putative reductase
putative glutaredoxin putativedehydratase/racemase
putativelipoprotein
putative dehydratase/racemase
no significant database hitss
Rhodanese-related sulfurtransferase
Rhodanese-related sulfurtransferase
SSU ribosomal protein S15p (S13e)
enoyl-CoA hydratase/isomerase family protein
CAIB/BAIF family protein
thiamine pyrophosphate enzyme
LysR family regulatory protein
Rhodanese-related sulfurtransferase
putative exported protein
No significant database matches
hypothetical protein
putative exported protein Acyl-CoA dehydrogenase
putative exported protein
putative regulatory protein
autotransporter



**Figura 15. A: Diagrama de Venn mostrando o compartilhamento de genes entre os 4 genomas gerados neste estudo e o genoma referência Tohama I. B: Diagrama de Venn mostrando o compartilhamento de genes entre os 4 genomas gerados neste estudo e o genoma referência Tohama I após a re-anotação**

### 4.3. Filogenômica e SNPs

Com o concatenado dos genes ortólogos (2050 genes) presentes em todos os 40 genomas da análise, foi construída uma árvore filogenética (Figura 16). Essa demonstrou-se não ser resolutive devido ao nível de conservação dos genomas analisados, ainda que seja possível distinguir a separação entre linhagens dos períodos pré e pós vacinais. Com o intuito de definirmos a relação mais direta entre os isolados, identificamos SNPs entre os genomas usados neste estudo e, a partir deste conjunto de informação, foi determinada a relação entre os isolados. Usando a plataforma *online SNPs finder* foram encontrados 32578 entre os 39 genomas de *B. pertussis* e o genoma de *B. bronchiseptica* RB50, dos quais 1675 eram exclusivos de *B. bronchiseptica*, posteriormente retirado das análises. Não foi encontrado nenhum SNP exclusivo dentre as linhagens brasileiras.

Os SNPs em regiões presentes em todos os isolados, um total de 826, foram usados para a construção da árvore de máxima verossimilhança.

Esta análise agrupou a maioria dos genomas pós vacinais, contendo o alelo 3 da região promotora do gene da toxina pertussis (*ptx*), em dois grupos A e B. As amostras do Brasil, todas do período pós vacinal, não pertencem a nenhum dos dois grupos A e B, apesar de duas delas (Bp100/RS/2008 e Bp168/PE/2013) manterem relação com estes grupos. Já, as outras duas amostras do Brasil (Bp1913/PA/2004 e Bp2041/PA/2004), apesar de apresentarem o alelo *ptxP3*, não pertencem a este agrupamento. Inclusive, a Bp2041, apresenta um perfil de MLST distinto das demais. Genomas do período pós vacinal apresentando o alelo 1 para o gene *ptxP* e genomas do período pré vacinal, formam dois outros grupos distintos (Figura 17).



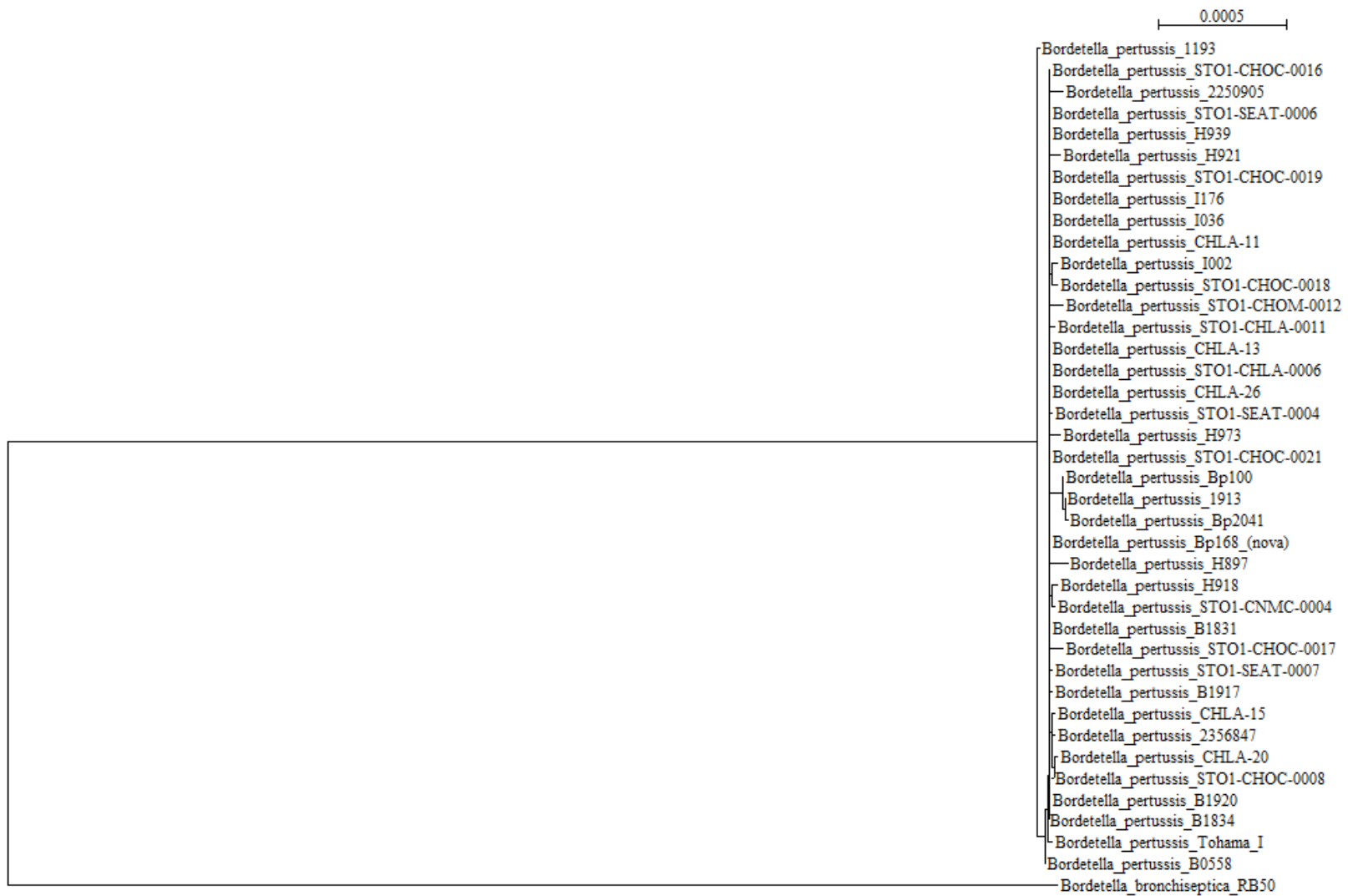
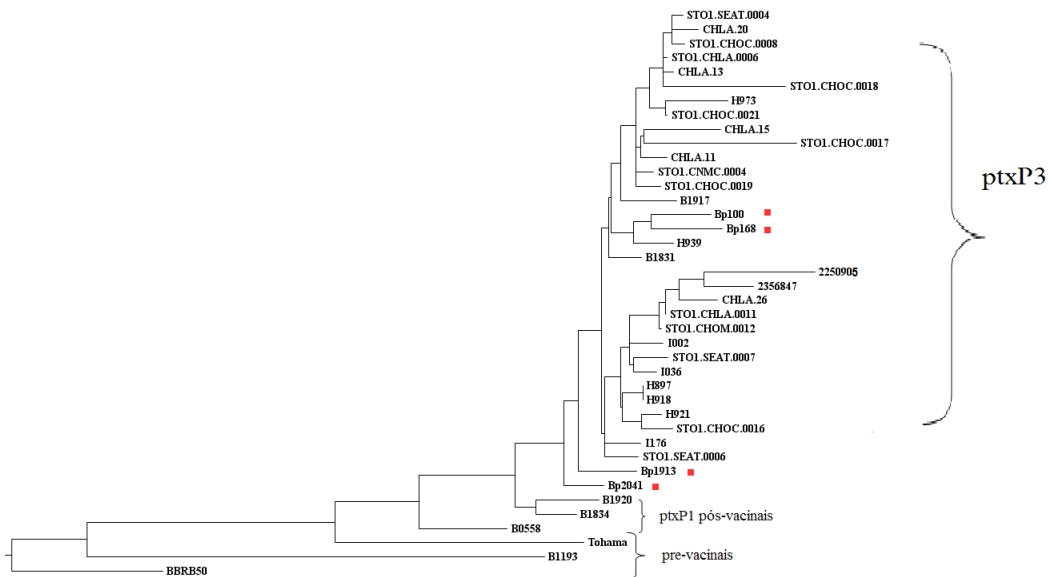


Figura 16. Relação filogenética entre *B. pertussis* considerando o conjunto de 2050 genes ortólogos.



**Figura 17- Árvore de máxima verossimilhança construída com 826 SNPs.** O quadrado vermelho indica os genomas gerados neste estudo. É possível observar a separação entre as linhagens pré e pós vacinal, e também a distinção entre as linhagens pós vacinais que se dividem em ptxP1 e ptxP3.

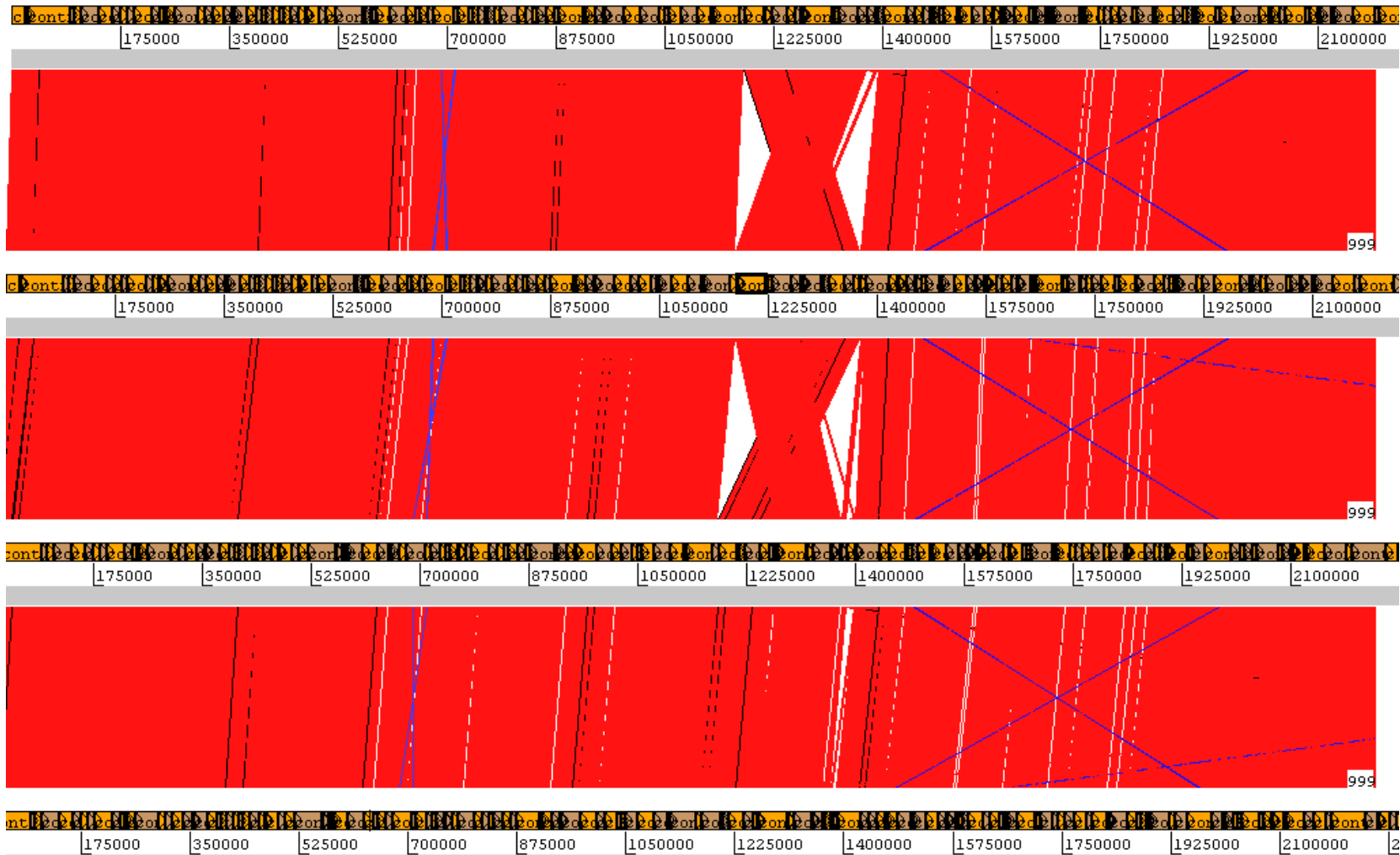
#### 4.4. Sintenia

O genoma de *B.pertussis* contém centenas de cópias de sequencias de inserção, particularmente a IS481. A presença destes elementos é um dos fatores para os frequentes rearranjos comuns nos genomas das *B. pertussis*.

Os *contigs* dos 4 genomas deste estudo foram ordenados com base no genoma de *B. broncispetica* RB50 e mapeados entre si por Blastn. Na visualização com o aplicativo ACT foi possível observar que a ordem gênica entre os 4 genomas estaria conservada, sendo assim, na maior parte dos casos, todo o conteúdo de um *contig* de um genoma está inteiramente contido em um outro. Ainda assim, foi possível observar que algumas regiões de um dado *contig* de um genoma possuem similaridade com múltiplas regiões de um

outro genoma, indicando um evento de duplicação gênica. Após verificação manual destas regiões, foi revelado que as mesmas correspondiam à sequência da IS481, que está presente parcialmente em vários dos *contigs* dos isolados sequenciados (Figura 18).

,



**Figura 18. Representação da sintenia entre os 4 genomas dos 4 isolados brasileiros gerados neste estudo.** As barras cinza indicam o genoma (de cima para baixo Bp100,Bp168,Bp1913,Bp2041), as faixas vermelhas indicam similaridade na mesma orientação, e as faixas azuis na orientação reversa. Os retângulos em rosa e laranja representam *contigs*. É possível sugerir uma ordem gênica conservada.

#### **4.5. Identificação de genes interrompidos pela IS481**

Sete regiões que contíguas no genoma de *B. pertussis* Tohama I apresentam-se interrompidas nos diferentes genomas deste estudo, algumas em apenas um dos genomas, outras nos 4 genomas gerados. Cinco dessas regiões interrompidas correspondem a segmentos intergênicos ou segmentos não transcritos adjacentes a códons de terminação, duas destas regiões correspondem a pseudogenes e uma corresponde ao gene regulador de transcrição da família MarR (Tabela 9).

**Tabela 9 .Posições interrompidas no genoma dos isolados do Brasil em relação ao genoma de referências Tohama I.** Na última coluna estão genomas que não apresentam IS481 nesta mesma região.

Nº	Isolado	Posição da Inserção em Tohama I	Locus	Gene/Regiões adjacentes	Ausente em
1	Bp168/Bp100/Bp1913/Bp2041	559016	Intergênico	3'-(1bp) -BP0552* 5'-(65bp)-BP0551+	B1193,B18323,CS
2	Bp168/Bp100/Bp1913/Bp2041	1015603	Intergênico	3'-(6bp)-BP0977+ 5'-(95bp)-BP0976*	2250905 ,H921,CHLA-26,B1193,B0558,CS
3	Bp168	1025764	BP0986	cusC(pseudogene)	Todos.
4	Bp168/Bp100/Bp1913	1518273	Intergênico	3'-(22bp)-BP1443+ 5'-(44bp)-BP1142+	STO1-SEAT-0006, CS,18323,B1193
5	Bp168/Bp100/Bp1913/Bp2041	2099174	BP1987	Regulador de transcrição da família MarR	B1193,B18323,CS
6	Bp168/Bp100//Bp1913/Bp2041	2764696	BP2611	Pseudogene	B1913,B18323,CS
7	Bp2041	2863994	BP2698	Endonuclease	Todos, exceto:  B0558, 225905

# Discussão

## 5. Discussão

Após mais de 50 anos de vacinação, pertússis voltou a ocorrer, mesmo em populações com alta cobertura vacinal, passando a ser a doença imunoprevenível mais frequente em países desenvolvidos (HE; MERTSOLA, 2008). Assim, como os antibióticos criam uma pressão, selecionando bactérias resistentes, a vacina também oferece uma pressão seletiva, que pode selecionar bactérias com variações que possibilitam sua evasão do sistema imune e, mesmo indivíduos vacinados, podem vir a manifestar sintomas associados a infecção (MOOI; VAN LOO; KING, 2001).

Duas versões de vacina para pertússis existem atualmente: a celular, que foi a primeira a ser aplicada e a acelular, que substituiu total ou parcialmente, dependendo do país, a vacina celular. Estudos realizados nos últimos anos caracterizaram, em alguns países, duas principais populações de *B. pertussis*, principal agente etiológico de pertússis, que são caracterizadas, principalmente, por polimorfismos em genes que codificam para as proteínas usadas na vacina acelular (CASSIDAY et al., 2000; VAN GENT et al., 2012a). Assim, isolados de *B. pertussis* são caracterizados, a partir do perfil deste conjunto de genes, como pré ou pós vacinais (BART et al., 2010, 2014; MAHARJAN et al., 2008)

A vacinação para pertússis faz parte do Programa Nacional de Imunização do Brasil (PNI) e para este fim é utilizada a vacina celular. A vacina acelular, também utilizada no país, está disponível em serviços privados, mas ainda assim, a re-emergência de pertússis é um fato no país (TREVIZAN; COUTINHO, 2008). O cenário epidemiológico desta infecção no país ainda não é claro, carecendo de várias informações, entre elas a caracterização genética de cepas *B. pertussis* circulantes, causadoras dos surtos da infecção no Brasil.

O presente estudo, que teve como principal objetivo revelar o genoma e realizar a caracterização genética de isolados de *B. pertussis* do Brasil, aplicou inicialmente a abordagem de MLST baseada em genes de virulência para definir linhagens e, a partir destas informações, selecionar os isolados que teriam os genomas sequenciados. A análise por MLST demonstrou que a maioria dos isolados do Brasil apresenta o perfil alélico *prn2-fim3B-ptxS1A-*



*ptxP3*. Este é o perfil prevalente em países que possuem alta cobertura vacinal como Argentina, Austrália, Canadá, Itália, Japão, Holanda, Suíça, Finlândia, Noruega, Israel, Estados Unidos (CAMPINS et al., 2013; HALLANDER et al., 2005; PARADOWSKA-STANKIEWICZ; RUDOWSKA, 2013; TAN; TRINDADE; SKOWRONSKI, 2005) e que corresponderia ao da população pós vacinal já que é distinto do perfil alélico de linhagens pré vacinais e cepas vacinais, incluindo a cepa usada para a produção da vacina que vem sendo utilizada no Brasil, a Bp137 (BOTTERO et al., 2012). As exceções são dois isolados que apresentaram o perfil *prn2-fim3A-ptxS1A-ptxP3*, ou seja, apresentam o alelo *fim3A* ao invés do alelo *fim3B*. Estes dois isolados são de Belém do Pará (ano de 2004) mas, naquele mesmo ano e lugar, também circulava o perfil dominante, ou seja, aquele que apresenta o alelo *fim3B*. Inferências Bayesianas indicam que *B. pertussis* apresentando *fim3B* surgiram entre 1986 e 1989 (BART et al., 2014) e apesar da diferença entre o produto destes dois alelos ser apenas um aminoácido, este aminoácido está presente em um epítipo de superfície que já foi demonstrado interagir com soro humano (WILLIAMSON; MATTHEWS, 1996) e, portanto, estaria supostamente sob seleção positiva.

Todos os isolados deste estudo albergavam o alelo *ptxP3*. Inferências Bayesianas demonstram que a mutação relativa a esse alelo surgiu em 1974 e 1977 (BART et al., 2014) e que a linhagem que apresenta esta mutação se espalhou globalmente desde então substituindo linhagens *ptxP1* existentes, porém é interessante notar que nenhum relato desse alelo foi feito no continente africano, onde a introdução da vacinação ainda é recente.

A região do alelo *ptxP3* corresponde a região promotora do *operon* para o gene da toxina pertússis e está associado a uma produção maior de toxina, quando comparada àquelas que ocorrem em isolados de *B. pertussis* apresentando os outros alelos de *ptxP* (MOOI et al., 2009). Portanto, além do aumento da produção do principal determinante da doença, a toxina pertússis, os isolados que possuem o alelo *ptxP3* também apresentam expressão aumentada de outros fatores de virulência (DE GOUW et al., 2014; KING et al., 2013), levando a hipótese de que estas *B. pertussis* poderiam ocasionar pertússis mesmo em indivíduos vacinados.

Além da hipótese da seleção da bactéria frente à vacinação, foi proposto que a introdução da vacina acelular teria ocasionado a re-emergência de

pertússis (ALLEN, 2013). De fato, estudos com babuínos demonstraram que a vacina celular gera proteção, enquanto a vacina acelular impediria o desenvolvimento da doença, mas não da infecção, gerando um número de indivíduos portadores assintomáticos, o que aumentaria a prevalência da *B. pertussis* nas populações (WARFEL; ZIMMERMAN; MERKEL, 2014).

É claro que a explicação para a re-emergência da pertússis não deve estar em uma única causa, além disso, sabemos que mesmo os países que tem programas de vacinação quer seja usando vacina celular e/ou acelular, apresentam falhas nos esquemas de vacinação, tornando mais complicada a definição das causas para a re-emergência de pertússis.

Portanto, com este projeto, pretendemos colaborar com informações genéticas completas em relação às *B. pertussis* causando pertússis no Brasil, que são informações absolutamente fundamentais mas não absolutas para o entendimento desta questão.

Até o presente momento, não se conhecia a(s) linhagem(s) de *B. pertussis* circulante(s) no Brasil. Sendo assim, o presente estudo lançou mão da genômica comparativa a fim de explorar a relação entre os isolados de *B. pertussis* do Brasil e destes com aqueles do resto do mundo que tinham seu genoma completamente sequenciado e publicado.

Dentre as espécies de patógenos considerados clonais *B. pertussis* é uma das que apresentam as menores taxas de substituição de nucleotídeos que é a de 1 SNP por 32Kb, que é significativamente menor que o da *Mycobacterium tuberculosis* que é 1 por 3 Kb e *Salmonella enterica* serovar Typhi com 1 por 10 kb, se aproximando da *Mycobacterium leprae*, 1/28 kb e *Yersinia pestis*, 1 por 35 kb (OCTAVIA et al., 2012). Portanto, não seria robusto realizar inferências evolutivas e epidemiológicas considerando apenas o conjunto de informações contidas em alguns Kbs do MLST sendo o conteúdo genômico global fundamental para o entendimento deste patógeno.

Até Julho de 2012, apenas oito genomas de *B. pertussis* estavam disponíveis no *GenBank* sendo seis deles, no formato *draft*, e todos os genomas de isolados contemporâneos eram de isolados da Holanda. Neste cenário selecionamos quatro isolados do Brasil, representando o tempo/espaco/perfil alélico, para recuperação da informação genômica global.

A tecnologia aplicada na recuperação dos genomas deste estudo foi a de sequenciamento de alto desempenho utilizando a plataforma GS 454 Junior. Sabemos que as distintas plataformas apresentam vantagens e desvantagens dependendo de características genéticas do organismo e do propósito do sequenciamento.

Em estudo que comparou diferentes tecnologias de sequenciamento e seu desempenho perante genomas com diferentes conteúdos GC, foi demonstrado que para as plataformas Pacific Biosciences, Ion Torrent Personal Genome Machine e Illumina MiSeq, o alto conteúdo GC de *B. pertussis* (~67%) não interfere na acurácia do sequenciamento por estas três tecnologias, e que possíveis erros estariam relacionados à tecnologia em si, e não ao conteúdo GC de *B. pertussis*. O presente estudo gerou o primeiro conjunto de informação genômica de *B. pertussis* a partir da tecnologia GS 454 Junior, e apesar da cobertura ter sido inferior àquelas obtidas por Quail e colaboradores (2012) (QUAIL et al., 2012), essa foi suficiente para recuperar todo o conteúdo genômico de nossas amostras já que *B. pertussis* apresenta um genoma de 4 Mb (LOMAN et al., 2012).

É sabido que esta tecnologia é robusta para sequenciamento de pequenos genomas, como o de bactérias, pois a cobertura obtida seria suficiente, além da vantagem do tamanho médio dos *reads* gerados ser torno de 450bp, o que compensaria uma “menor cobertura”. Outra particularidade do genoma de *B. pertussis* é a presença de um elevado número de sequências repetitivas, maiores que 1 Kb, que correspondem a IS481. Logo, uma tecnologia que apresente *reads* mais longos é de extrema utilidade para este contexto genômico.

Comparando as métricas obtidas no nosso estudo com àquelas de Quail e colaboradores (2012) nota-se que o tamanho médio dos *reads* gerados ao se sequenciar *B. pertussis* na plataforma GS 454 Junior é maior que as plataformas Ion Torrent Personal Genome Machine (115 pb) e Illumina MiSeq (151 pb), sendo inferior apenas ao Pacific Biosciences (1.1Kb), mas essa plataforma foi a que demonstrou menor acurácia dentre as testadas. Sendo assim, consideramos que a plataforma GS 454 Junior é, e foi, adequada para o sequenciamento de *B. pertussis*.

O número de *contigs* obtidos com a montagem, que variou de 273 a 277, é compatível com o esperado, já que como mencionado, *B. pertussis* apresenta inúmeras cópias da sequência de inserção IS481 (~ 1Kb), o que considerando *reads* de aproximadamente 450 pb torna a montagem limitada. Tanto que, a maioria destes *contig* está flanqueada por segmentos da IS481 e o número de *contigs* gerados após a montagem se aproxima ao número de cópias desta IS identificados nos genomas montados e inferidos em alguns trabalhos (BOUCHEZ et al., 2008).

Como demonstrado por Stibitz e Yang (1999) e mais recentemente por Harvill e colaboradores (2013), uma característica do genoma de *B. pertussis* é a dinâmica da ordem gênica, já que é observado, em distintos isolados, rearranjos de conjunto de genes, geralmente flanqueados por sequências correspondentes a ISs. Este processo seria consequência de eventos de recombinação homóloga entre as numerosas cópias da sequência de inserção IS481, que acaba por mover toda a região contida entre elas (PRESTON; PARKHILL; MASKELL, 2004).

Em nosso estudo não foi possível avaliar a presença de rearranjos mediados pela IS481, pois pela natureza do sequenciamento, não tínhamos informação que permitisse esta inferência. Portanto, procuramos por rearranjos não dependentes da IS481, que não foram observados entre os genomas dos isolados de *B. pertussis* do Brasil.

Considerando a importância da IS481 na evolução de *B. pertussis* (HEIKKINEN et al., 2007b; PARK et al., 2012a; PRESTON; PARKHILL; MASKELL, 2004) desenvolvemos uma metodologia que revela as posições deste elemento nos genomas dos isolados do Brasil, e que são distintas no genoma referência Tohama I. A metodologia aplicada utilizou *reads* provindos da plataforma GS 454 Junior, porém é possível que a mesma seja extrapolada para diferentes plataformas, e talvez diferentes sequências repetitivas.

Num total foram identificadas sete regiões nas quais a IS481 está inserida. Duas dessas regiões correspondem a pseudogenes, e três são regiões intergênicas. A inserção dessa sequência se é sítio-específica, ocorrendo quando a transposase codificada pela IS encontra a sequência CGTA (VAN DER ZEE et al., 1996) presente em inúmeras regiões do genoma. Porém, caso a inserção deste elemento interrompa genes essenciais a

sobrevivência da bactéria seja comprometida, dessa maneira é esperado que esta sequência seja encontrada principalmente em regiões intergênicas, genes acessórios, pseudogenes, entre outras regiões que não desempenham papel fundamental na sobrevivência dos organismos. Porém, caso a inativação de um determinado gene confira uma vantagem, é possível que essa inserção se fixe, mesmo em genes estruturais como tem sido observado no gene da pertactina. Pertactina é um antígeno de *B. pertussis* que é um dos componentes das vacinas acelulares, sua perda pode resultar no escape do mutante da resposta imune do hospedeiro (LAM et al., 2014; OTSUKA et al., 2012; PAWLOSKI et al., 2014).

As duas outras regiões com IS481, correspondem ao gene de uma endonuclease e a um gene relacionado a regulação de transcrição da família MarR. Outras linhagens pós vacinais, apresentam IS481 nessas mesmas regiões, o que poderia indicar uma vantagem frente a pressão da vacina especialmente, quando consideramos o gene regulador, já que este está envolvido em regulação positiva/negativa em vias relacionadas à virulência em algumas espécies (ELLISON; MILLER, 2006).

As sete regiões identificadas como interrompidas nos genomas de isolados do Brasil foram comparadas com os outros genomas disponíveis no *Genbank* (Tabela 9). Quando relacionamos estes dados com a árvore de SNPs gerada neste estudo, observamos que, de maneira geral, uma vez que o evento de inserção ocorreu, o mesmo foi mantido em todos os grupos descendentes. Porém as regiões 2, 4 e 7 esta relação não foi observada em todos os genomas pertencentes ao mesmo grupo filogenético. Propomos que a explicação para essa aparente incongruência pode ser computacional ou biológica. Por se tratar de um elemento móvel, o mesmo pode ter sido excisado e a região interrompida foi refeita. Outra hipótese está no fato de que este evento de inserção tenha ocorrido por convergência evolutiva. Outra possibilidade está no fato de que estamos tratando de dados computacionais e tanto a montagem desses genomas quanto a metodologia empregada poderia ter resultado em vieses.

A inserção na região 3 está presente apenas no isolado Bp168 (PE, 2013), que é o genoma do isolado mais atual deste estudo, seria interessante

observar em novos isolados a presença da IS481, o que pode vir a revelar novas informações a respeito da epidemiologia de pertússis no Brasil.

Uma de nossas análises foi o estabelecimento do genoma core e pangenoma de *B. pertussis*. Estes conjuntos de genes foram definidos a partir da informação contida em 17 genomas representativos no tempo e espaço já que consideramos genomas de isolados obtidos num intervalo de aproximadamente 60 anos, em três continentes. O genoma core e pangenoma deste universo de isolados é representado por 2238 e 4416 genes, respectivamente.

A curva de crescimento do pangenoma apresenta duas inclinações positivas, uma após a adição do segundo genoma e outra após a adição do oitavo genoma. A explicação para o primeiro incremento e estabilização pode ser pelo fato de todos estes genomas serem de isolados da Holanda, independentemente se do período pré ou pós vacinal, da mesma forma, o segundo incremento ocorre ao ser introduzido os primeiros genomas dos isolados dos Estados Unidos, todos pós-vacinal, sendo assim, o espaço parece estar sendo um dos determinantes para a expansão do pangenoma.

Uma outra explicação pode estar no fato de que os genomas provenientes de isolados da Holanda (do segundo ao sétimo) foram anotados pelo RAST, sendo assim, haveria um viés de anotação, já que, o conteúdo gênico do genoma Tohama I (primeiro genoma da análise) foi manualmente anotado e a anotação pelo RAST normalmente identifica um número maior de genes. Da mesma forma, o conteúdo gênico dos isolados dos Estados Unidos foram determinados por um outro sistema de anotação que pode ter seus vieses. De qualquer forma, estes dois fatores combinados explicariam as acentuações observadas na curva.

A curva da análise de genoma core demonstra uma queda contínua, independente do tempo-espaço de obtenção do isolado, indicando a contínua perda e/ou inativação de genes pela espécie que estaria ainda em processo de adaptação ao hospedeiro.

Ainda assim, é preciso saber que estes não são resultados definitivos, uma vez que muitos dos genomas utilizados em nosso estudo possuem montagens e anotações diferentes entre si, enquanto que, os genomas utilizados no estudo de Park e colaboradores (2012) estão completos em um

único *contig*, foram manualmente anotados mas, por serem apenas três, não seriam representativos da diversidade de *B. pertussis*.

Dos 4548 *cluster* gênicos identificados pelo OrthoMCL, àqueles presentes em apenas um genoma eram considerados candidatos à genes únicos. A ausência de genes únicos dos isolados do Brasil indica que eventos de aquisição gênica não ocorreram nas linhagens circulantes no país, compatíveis com dados da literatura (PARK et al., 2012a; PARKHILL et al., 2003). Sendo assim, o conteúdo gênico de isolados Brasileiros foi comparado ao genoma referência Tohama I. Genes identificados como únicos ao terem sua sequência de aminoácidos mapeada contra o proteoma do genoma de Tohama I, de fato não apresentaram nenhum *hit*, porém ao mapear-se sua sequência de nucleotídica ao genoma de Tohama I, muitos deles obtinham identidade de 100%. Isso significa que apesar de ambos os genomas possuírem determinada sequência codificadora, em apenas um deles ela foi identificada pelo sistema de anotação de escolha, o que gerou falsos genes únicos. Este fato é mais evidente quando comparamos o número de genes únicos dos isolados do Brasil em relação a Tohama I usando sua anotação original, que foi de 530, com o número de genes únicos usando a anotação feita pelo RAST (AZIZ et al., 2008), mesmo anotador usado para o genoma de isolados do Brasil, que foi de 133. Dos 44 genes de fato únicos, grande parte constituía de genes hipotéticos, porém é interessante notar que alguns estavam relacionados a regulação gênica, o que poderia explicar uma maior virulência e adaptabilidade de isolados pós-vacinais (DE GOUW et al., 2014; KING et al., 2013; MOOI, 2010).

A recuperação da informação genômica de isolados do Brasil possibilitou a análise *in silico* da presença de alelos relacionados à resistência a alguns antibióticos. Este tipo de avaliação é bastante útil principalmente no caso de bactérias fastidiosas como a *B. pertussis*, especialmente quando vem sido descrita a emergência de resistência associada a variações em genes do genoma core, como a mutação A2047G no domínio V do gene que codifica a subunidade ribossômica 23S, associada a resistência a macrolídeos como eritromicina, claritromicina, e azitromicina, detectada em isolados da China com o perfil alélico *ptxS1A/ptxP1* (WANG et al., 2014) e com o perfil *ptxA/ptxP3* em isolados da França (GUILLOT et al., 2012). Assim como no gene *gyrA* (D87G)

que leva a resistência a quinolonas, como ácido nalidíxico, observada em isolados do Japão (OHTSUKA et al., 2009). Nenhum dos polimorfismos em genes do genoma core associados a resistência a esses antibióticos, assim como polimorfismos nos genes *parC* e *E*, também associados a resistência a quinolonas, foi observado entre nossos isolados, e/ou genomas disponíveis no *GenBank*.

Nosso estudo primeiro identificou genes ortólogos entre os 4 isolados do Brasil, 6 isolados da Holanda, 28 isolados dos Estados Unidos e o genoma referência Tohama I. O genoma referência de *B. bronchiseptica* RB50 também foi incluído nas análises. Os genes do genoma core, foram usados para o estabelecimento da relação filogenética dos 4 isolados do Brasil e *B. pertussis*. Apesar de repetir a topologia obtida por Bart e colaboradores (2010) (BART et al., 2010), as diferenças entre os genomas, que é em torno de 1%, fica “diluída” quando é usado o conteúdo global. A fim de valorizar as diferenças, foi estabelecida uma relação filogenética baseada, unicamente, nos SNPs.

A árvore baseada em SNPs apresenta uma topologia semelhante àquela obtida com os genes ortólogos concatenados, quando consideramos o agrupamento entre as linhagens pre e pós vacinais, porém é visualmente mais resolutive. Dois principais clusters foram definidos, um com isolados pré-vacinais e o outro com os isolados pós-vacinais que apresentava um sub-cluster com os isolados caracterizados pelo alelo 3 para o gene *ptx*.

É interessante observar que, aparentemente, a topologia da árvore baseada em SNPs repete a topologia demonstrada pela árvore baseada em MLST dos genes de virulência. Isso também foi observado por Bart e colaboradores (2014) (BART et al., 2014) com um número maior de genomas representativos no tempo e espaço.

Sendo assim, a maioria dos isolados com o mesmo perfil alélico estariam agrupados em uma árvore baseada em SNPs. Isto porém não é demonstrado para o isolado Bp1913 (PA, 2004, perfil *prn3-ptxS1A-ptxP3-fim3B*) que, na árvore baseada em SNPs, está fora do grupo de isolados com o seu perfil alélico.

Na inferência filogenética baseada nos SNPs, dois isolados, BP2041 e BP1913 – ambos de Belém do Pará, 2004 - não estão agrupados com os demais isolados do Brasil e com pós-vacinais da Europa e Estados Unidos



(todos *ptxP3*). Estes isolados podem fazer parte de subpopulações de *B. pertussis* que co-circulam com a linhagem pandêmica, principalmente em regiões onde a cobertura vacinal não foi absoluta, isto já foi visto em outros países (SCHMIDTKE et al., 2012; SHUEL et al., 2013). De fato, a posição do BP2041 (que é o único com o alelo *fim3A*), e que teria o genoma mais divergente do grupo, seguido do BP1913 (que já teria todo o perfil do grupo principal pós-vacinal), é uma evidência para a existência destas subpopulações.

O genoma da linhagem usada para a produção da vacina no Brasil (Bp173) (PEREIRA et al., 2005) ainda está sendo sequenciado (<http://www.ncbi.nlm.nih.gov/bioproject/221122>) e por isso não foi incluído em nossas análises. Ainda assim, a partir de seu perfil alélico (*prn1-ptxS1D-ptxP2-fim3A*) (BOTTERO et al., 2012), é possível inferir que seja uma linhagem distinta da prevalente atualmente circulante no país e que pode, por suas características antigênicas, não estar mais sendo eficiente frente a atual *B. pertussis* causadora de pertússis no país.

# Conclusões

## 6. Conclusões

- Há, pelo menos, duas linhagens de *B. pertussis* com distintos perfis de genes de virulência identificados no Brasil, no qual o perfil circulante é o prevalente no país e em outros países com alta cobertura vacinal;
- A plataforma GS 454 Junior mostrou-se eficiente para o sequenciamento de um genoma com as características do genoma de *B. pertussis*;
- Os surtos recentes de pertussis no RS e PE são determinados pela linhagem pandêmica de *B. pertussis*;
- A cepa usada para a produção da vacina no Brasil é distinta antigenicamente das linhagens de *B. pertussis* circulantes no país.

## **Referências Bibliográficas**

## 7. Referências Bibliográficas

- ALLEN, A. Public health. The pertussis paradox. **Science (New York, N.Y.)**, v. 341, n. 6145, p. 454–5, 2 ago. 2013.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–10, 5 out. 1990.
- ARICÒ, B. et al. Structural and genetic analysis of the bvg locus in Bordetella species. **Molecular microbiology**, v. 5, n. 10, p. 2481–91, out. 1991.
- AZIZ, R. K. et al. The RAST Server: rapid annotations using subsystems technology. **BMC genomics**, v. 9, p. 75, jan. 2008.
- BAKER, J. P.; KATZ, S. L. Childhood vaccine development: an overview. **Pediatric research**, v. 55, n. 2, p. 347–56, fev. 2004.
- BART, M. J. et al. Comparative genomics of prevaccination and modern Bordetella pertussis strains. **BMC genomics**, v. 11, p. 627, jan. 2010.
- BART, M. J. et al. Global Population Structure and Evolution of Bordetella pertussis and Their Relationship with Vaccination. **mBio**, v. 5, n. 2, jan. 2014.
- BLANKENBERG, D. et al. Manipulation of FASTQ data with Galaxy. **Bioinformatics (Oxford, England)**, v. 26, n. 14, p. 1783–5, 15 jul. 2010.
- BORDET, J.; OCTAVE, G. Le microbe de la coqueluche. **Ann. Inst. Pasteur**, v. 20, p. 731–741, 1906.
- BORISOVA, O. et al. Antigenic divergence between Bordetella pertussis clinical isolates from Moscow, Russia, and vaccine strains. **Clinical and vaccine immunology : CVI**, v. 14, n. 3, p. 234–8, mar. 2007.
- BOTTERO, D. et al. Genotypic and phenotypic characterization of Bordetella pertussis strains used in different vaccine formulations in Latin America. **Journal of applied microbiology**, v. 112, n. 6, p. 1266–76, jun. 2012.
- BOUCHEZ, V. et al. Genomic content of Bordetella pertussis clinical isolates circulating in areas of intensive children vaccination. **PloS one**, v. 3, n. 6, p. e2437, jan. 2008.
- CAMPINS, M. et al. [Whooping cough in Spain. Current epidemiology, prevention and control strategies. Recommendations by the Pertussis Working

- Group]. **Enfermedades infecciosas y microbiología clínica**, v. 31, n. 4, p. 240–53, abr. 2013.
- CARVER, T. J. et al. ACT: the Artemis Comparison Tool. **Bioinformatics (Oxford, England)**, v. 21, n. 16, p. 3422–3, 15 ago. 2005.
- CASSIDAY, P. et al. Polymorphism in Bordetella pertussis pertactin and pertussis toxin virulence factors in the United States, 1935-1999. **The Journal of infectious diseases**, v. 182, n. 5, p. 1402–8, nov. 2000.
- CHARLES, I. G. et al. Molecular cloning and characterization of protective outer membrane protein P.69 from Bordetella pertussis. **Proceedings of the National Academy of Sciences of the United States of America**, v. 86, n. 10, p. 3554–8, maio 1989.
- CHARLES, I. G. et al. Identification and characterization of a protective immunodominant B cell epitope of pertactin (P.69) from Bordetella pertussis. **European journal of immunology**, v. 21, n. 5, p. 1147–53, maio 1991.
- CHIAPPINI, E. et al. Pertussis re-emergence in the post-vaccination era. **BMC infectious diseases**, v. 13, n. 1, p. 151, jan. 2013.
- CONE, T. C. Whooping cough is first described as a disease sui generis by Baillou in 1640. **Pediatrics**, v. 46, n. 4, p. 522, out. 1970.
- CONTRERAS-MOREIRA, B.; VINUESA, P. GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. **Applied and environmental microbiology**, v. 79, n. 24, p. 7696–701, dez. 2013.
- CUMMINGS, C. A. et al. Bordetella species are distinguished by patterns of substantial gene loss and host adaptation. **Journal of bacteriology**, v. 186, n. 5, p. 1484–92, mar. 2004.
- DE GOUW, D. et al. Differentially expressed genes in Bordetella pertussis strains belonging to a lineage which recently spread globally. **PloS one**, v. 9, n. 1, p. e84523, jan. 2014.
- DE MAGISTRIS, M. T. et al. Human T cell clones define S1 subunit as the most immunogenic moiety of pertussis toxin and determine its epitope map. **The Journal of experimental medicine**, v. 169, n. 5, p. 1519–32, 1 maio 1989.
- DIAVATOPOULOS, D. A. et al. Characterization of a highly conserved island in the otherwise divergent Bordetella holmesii and Bordetella pertussis genomes. **Journal of bacteriology**, v. 188, n. 24, p. 8385–94, dez. 2006.
- EDWARDS, K. M.; HALASA, N. B. Commentary: is pertussis disease increasing? **International journal of epidemiology**, v. 33, n. 2, p. 365–6, abr. 2004.

ELLISON, D. W.; MILLER, V. L. Regulation of virulence by members of the MarR/SlyA family. **Current opinion in microbiology**, v. 9, n. 2, p. 153–9, abr. 2006.

FIETT, J. et al. The new strategy for allele identification of the genes coding for pertussis toxin subunit S1 (ptx S1) and pertactin (prn) in *Bordetella pertussis*. **Journal of microbiological methods**, v. 55, n. 3, p. 651–66, dez. 2003.

FINN, T. M.; STEVENS, L. A. Tracheal colonization factor: a *Bordetella pertussis* secreted virulence determinant. **Molecular microbiology**, v. 16, n. 4, p. 625–34, maio 1995.

FU, L. et al. CD-HIT: accelerated for clustering the next-generation sequencing data. **Bioinformatics (Oxford, England)**, v. 28, n. 23, p. 3150–2, 1 dez. 2012.

GUILLOT, S. et al. Macrolide-resistant *Bordetella pertussis* infection in newborn girl, France. **Emerging infectious diseases**, v. 18, n. 6, p. 966–8, jun. 2012.

GUINDON, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. **Systematic biology**, v. 59, n. 3, p. 307–21, maio 2010.

HALL., T. . BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows. **Nucl. Acids Symp**, v. 41, p. 95–98, 1999.

HALLANDER, H. et al. *Bordetella pertussis* strains circulating in Europe in 1999 to 2004 as determined by pulsed-field gel electrophoresis. **Journal of clinical microbiology**, v. 45, n. 10, p. 3257–62, out. 2007.

HALLANDER, H. O. et al. Shifts of *Bordetella pertussis* variants in Sweden from 1970 to 2003, during three periods marked by different vaccination programs. **Journal of clinical microbiology**, v. 43, n. 6, p. 2856–65, jun. 2005.

HARVILL, E. T. et al. Genome Sequences of 28 *Bordetella pertussis* U.S. Outbreak Strains Dating from 2010 to 2012. **Genome announcements**, v. 1, n. 6, jan. 2013.

HE, Q.; MERTSOLA, J. Factors contributing to pertussis resurgence. **Future microbiology**, v. 3, n. 3, p. 329–39, jun. 2008.

HEGERLE, N.; GUIISO, N. Epidemiology of whooping cough & typing of *Bordetella pertussis*. **Future microbiology**, v. 8, n. 11, p. 1391–403, nov. 2013.

HEIKKINEN, E. et al. Comparative genomics of *Bordetella pertussis* reveals progressive gene loss in Finnish strains. **PloS one**, v. 2, n. 9, p. e904, jan. 2007a.

HEIKKINEN, E. et al. Comparative genomics of *Bordetella pertussis* reveals progressive gene loss in Finnish strains. **PloS one**, v. 2, n. 9, p. e904, 19 jan. 2007b.

HEWLETT, E. L. et al. Extracytoplasmic adenylate cyclase of *Bordetella pertussis*. **Proceedings of the National Academy of Sciences of the United States of America**, v. 73, n. 6, p. 1926–30, jun. 1976.

HUANG, W. et al. ART: a next-generation sequencing read simulator. **Bioinformatics (Oxford, England)**, v. 28, n. 4, p. 593–4, 15 fev. 2012.

KAMACHI, K. et al. Plasmid pBP136 from *Bordetella pertussis* represents an ancestral form of IncP-1beta plasmids without accessory mobile elements. **Microbiology (Reading, England)**, v. 152, n. Pt 12, p. 3477–84, dez. 2006.

KATOH, K.; TOH, H. Recent developments in the MAFFT multiple sequence alignment program. **Briefings in bioinformatics**, v. 9, n. 4, p. 286–98, jul. 2008.

KENDRICK, P.; ELDERING, G.; HEALTH, M. D. OF. **A Study in Active Immunization Against Pertussis**. [s.l.: s.n.].

KIMURA, M.; OTA, T. On the stochastic model for estimation of mutational distance between homologous proteins. **Journal of molecular evolution**, v. 2, n. 1, p. 87–90, 29 dez. 1972.

KING, A. J. et al. Role of the polymorphic region 1 of the *Bordetella pertussis* protein pertactin in immunity. **Microbiology (Reading, England)**, v. 147, n. Pt 11, p. 2885–95, nov. 2001.

KING, A. J. et al. Genome-Wide Gene Expression Analysis of *Bordetella pertussis* Isolates Associated with a Resurgence in Pertussis: Elucidation of Factors Involved in the Increased Fitness of Epidemic Strains. **PLoS one**, v. 8, n. 6, p. e66150, jan. 2013.

LAM, C. et al. Rapid increase in pertactin-deficient *Bordetella pertussis* isolates, Australia. **Emerging infectious diseases**, v. 20, n. 4, p. 626–33, abr. 2014.

LI, H.; DURBIN, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. **Bioinformatics (Oxford, England)**, v. 26, n. 5, p. 589–95, 1 mar. 2010.

LI, L.; STOECKERT, C. J.; ROOS, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. **Genome research**, v. 13, n. 9, p. 2178–89, 1 set. 2003.

LINNEMANN, C. C. et al. Use of pertussis vaccine in an epidemic involving hospital staff. **Lancet**, v. 2, n. 7934, p. 540–3, 20 set. 1975.

LOCHT, C. et al. The filamentous haemagglutinin, a multifaceted adhesion produced by virulent *Bordetella* spp. **Molecular microbiology**, v. 9, n. 4, p. 653–60, ago. 1993.



LOCHT, C.; KEITH, J. M. Pertussis toxin gene: nucleotide sequence and genetic organization. **Science (New York, N.Y.)**, v. 232, n. 4755, p. 1258–64, 6 jun. 1986.

LOMAN, N. J. et al. Performance comparison of benchtop high-throughput sequencing platforms. **Nature biotechnology**, v. 30, n. 5, p. 434–9, maio 2012.

MAHARJAN, R. P. et al. Genome-wide analysis of single nucleotide polymorphisms in *Bordetella pertussis* using comparative genomic sequencing. **Research in microbiology**, v. 159, n. 9-10, p. 602–8, 2008.

MATTOO, S.; CHERRY, J. D. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to *Bordetella pertussis* and other *Bordetella* subspecies. **Clinical microbiology reviews**, v. 18, n. 2, p. 326–82, abr. 2005.

MERKEL, T. J. et al. Analysis of *bvgR* expression in *Bordetella pertussis*. **Journal of bacteriology**, v. 185, n. 23, p. 6902–12, dez. 2003.

MOOI, F. R. et al. Polymorphism in the *Bordetella pertussis* virulence factors P.69/pertactin and pertussis toxin in The Netherlands: temporal trends and evidence for vaccine-driven evolution. **Infection and immunity**, v. 66, n. 2, p. 670–5, mar. 1998.

MOOI, F. R. et al. *Bordetella pertussis* strains with increased toxin production associated with pertussis resurgence. **Emerging infectious diseases**, v. 15, n. 8, p. 1206–13, ago. 2009.

MOOI, F. R. *Bordetella pertussis* and vaccination: the persistence of a genetically monomorphic pathogen. **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 10, n. 1, p. 36–49, jan. 2010.

MOOI, F. R.; HE, Q.; GUIZO, N. ***Bordetella: Molecular Microbiology***. [s.l.] Horizon Scientific Press, 2007. p. 18–9

MOOI, F. R.; VAN DER MAAS, N. A. T.; DE MELKER, H. E. Pertussis resurgence: waning immunity and pathogen adaptation – two sides of the same coin. **Epidemiology and Infection**, p. 1–10, 13 fev. 2013a.

MOOI, F. R.; VAN DER MAAS, N. A. T.; DE MELKER, H. E. Pertussis resurgence: waning immunity and pathogen adaptation - two sides of the same coin. **Epidemiology and infection**, p. 1–10, 13 fev. 2013b.

MOOI, F. R.; VAN LOO, I. H.; KING, A. J. Adaptation of *Bordetella pertussis* to vaccination: a cause for its reemergence? **Emerging infectious diseases**, v. 7, n. 3 Suppl, p. 526–8, jan. 2001.

MUSSER, J. M. et al. Genetic diversity and relationships in populations of *Bordetella* spp. **Journal of bacteriology**, v. 166, n. 1, p. 230–7, abr. 1986.

OCTAVIA, S. et al. Newly emerging clones of *Bordetella pertussis* carrying prn2 and ptxP3 alleles implicated in Australian pertussis epidemic in 2008-2010. **The Journal of infectious diseases**, v. 205, n. 8, p. 1220–4, 15 abr. 2012.

OHTSUKA, M. et al. Emergence of quinolone-resistant *Bordetella pertussis* in Japan. **Antimicrobial agents and chemotherapy**, v. 53, n. 7, p. 3147–9, jul. 2009.

OKONECHNIKOV, K.; GOLOSOVA, O.; FURSOV, M. Unipro UGENE: a unified bioinformatics toolkit. **Bioinformatics (Oxford, England)**, v. 28, n. 8, p. 1166–7, 15 abr. 2012.

OTSUKA, N. et al. Prevalence and genetic characterization of pertactin-deficient *Bordetella pertussis* in Japan. **PloS one**, v. 7, n. 2, p. e31985, jan. 2012.

PARADOWSKA-STANKIEWICZ, I.; RUDOWSKA, J. Pertussis in Poland in 2011. **Przegląd epidemiologiczny**, v. 67, n. 2, p. 199–201, 319–21, jan. 2013.

PARK, J. et al. Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. **BMC genomics**, v. 13, p. 545, jan. 2012a.

PARK, J. et al. Comparative genomics of the classical *Bordetella* subspecies: the evolution and exchange of virulence-associated diversity amongst closely related pathogens. **BMC genomics**, v. 13, p. 545, jan. 2012b.

PARKHILL, J. et al. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. **Nature genetics**, v. 35, n. 1, p. 32–40, set. 2003.

PAWLOSKI, L. C. et al. Prevalence and molecular characterization of pertactin-deficient *Bordetella pertussis* in the United States. **Clinical and vaccine immunology : CVI**, v. 21, n. 2, p. 119–25, fev. 2014.

PEREIRA, A. et al. Comparative analysis of a *Bordetella pertussis* patient isolated strain and classical strains used in the pertussis vaccine. **Vaccine**, v. 23, n. 34, p. 4353–8, 29 jul. 2005.

POYNTEN, M. et al. Temporal trends in circulating *Bordetella pertussis* strains in Australia. **Epidemiology and infection**, v. 132, n. 2, p. 185–93, abr. 2004.

PRESTON, A.; PARKHILL, J.; MASKELL, D. J. The bordetellae: lessons from genomics. **Nature reviews. Microbiology**, v. 2, n. 5, p. 379–90, maio 2004.

QUAIL, M. A. et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. **BMC genomics**, v. 13, p. 341, jan. 2012.

QUEENAN, A. M.; CASSIDAY, P. K.; EVANGELISTA, A. Pertactin-negative variants of *Bordetella pertussis* in the United States. **The New England journal of medicine**, v. 368, n. 6, p. 583–4, 7 fev. 2013.

R CORE TEAM. **R: A Language and Environment for Statistical Computing** Viena, Áustria, 2013. Disponível em: <<http://www.r-project.org/>>

RESP. **INFORME TÉCNICO COQUELUCHE 2011**. Disponível em: <[http://www.cve.saude.sp.gov.br/html/resp/pdf/IF11\\_COQUELUCHE.pdf](http://www.cve.saude.sp.gov.br/html/resp/pdf/IF11_COQUELUCHE.pdf)>. Acesso em: 25 jun. 2014.

REVIEWS, C. T. **e-Study Guide for: Nurse Practitioner Certification Examination and Practice Preparation: Nursing, Nursing (Google eBook)**. [s.l.] Cram101 Textbook Reviews, 2012. p. 112

RISSMAN, A. I. et al. Reordering contigs of draft genomes using the Mauve aligner. **Bioinformatics (Oxford, England)**, v. 25, n. 16, p. 2071–3, 15 ago. 2009.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol. Biol. Evol.**, v. 4, n. 4, p. 406–425, 1 jul. 1987.

SATO, Y.; KIMURA, M.; FUKUMI, H. Development of a pertussis component vaccine in Japan. **Lancet**, v. 1, n. 8369, p. 122–6, 21 jan. 1984.

SCHMIDTKE, A. J. et al. Population diversity among *Bordetella pertussis* isolates, United States, 1935-2009. **Emerging infectious diseases**, v. 18, n. 8, p. 1248–55, ago. 2012.

SHUEL, M. et al. Genetic analysis of *Bordetella pertussis* in Ontario, Canada reveals one predominant clone. **International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases**, v. 17, n. 6, p. e413–7, jun. 2013.

SONG, J. et al. SNPsFinder--a web-based application for genome-wide discovery of single nucleotide polymorphisms in microbial genomes. **Bioinformatics (Oxford, England)**, v. 21, n. 9, p. 2083–4, 1 maio 2005.

STIBITZ, S.; YANG, M. S. Genomic plasticity in natural populations of *Bordetella pertussis*. **Journal of bacteriology**, v. 181, n. 17, p. 5512–5, set. 1999.

SWINDELL, S. R.; PLASTERER, T. N. SEQMAN. Contig assembly. **Methods in molecular biology (Clifton, N.J.)**, v. 70, p. 75–89, jan. 1997.

TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular biology and evolution**, v. 28, n. 10, p. 2731–9, out. 2011.

TAN, T.; TRINDADE, E.; SKOWRONSKI, D. Epidemiology of pertussis. **The Pediatric infectious disease journal**, v. 24, n. 5 Suppl, p. S10–8, maio 2005.

TARANGER, J. et al. Mass vaccination of children with pertussis toxoid--decreased incidence in both vaccinated and nonvaccinated persons. **Clinical infectious diseases : an official publication of the Infectious Diseases Society of America**, v. 33, n. 7, p. 1004–10, 1 out. 2001.

TREVIZAN, S.; COUTINHO, S. E. D. [Epidemiological profile of pertussis in Rio Grande do Sul State, Brazil: a study of the correlation between incidence and vaccine coverage]. **Cadernos de saúde pública / Ministério da Saúde, Fundação Oswaldo Cruz, Escola Nacional de Saúde Pública**, v. 24, n. 1, p. 93–102, jan. 2008.

TSANG, R. S. W. et al. Polymorphisms of the fimbria fim3 gene of Bordetella pertussis strains isolated in Canada. **Journal of clinical microbiology**, v. 42, n. 11, p. 5364–7, nov. 2004.

VAN DER ZEE, A. et al. The differentiation of Bordetella parapertussis and Bordetella bronchiseptica from humans and animals as determined by DNA polymorphism mediated by two different insertion sequence elements suggests their phylogenetic relationship. **International journal of systematic bacteriology**, v. 46, n. 3, p. 640–7, jul. 1996.

VAN GENT, M. et al. Studies on Prn variation in the mouse model and comparison with epidemiological data. **PloS one**, v. 6, n. 3, p. e18014, jan. 2011.

VAN GENT, M. et al. Small mutations in Bordetella pertussis are associated with selective sweeps. **PloS one**, v. 7, n. 9, p. e46407, jan. 2012a.

VAN GENT, M. et al. Small mutations in Bordetella pertussis are associated with selective sweeps. **PloS one**, v. 7, n. 9, p. e46407, jan. 2012b.

VAN LOO, I. H. M. et al. Multilocus sequence typing of Bordetella pertussis based on surface protein genes. **Journal of clinical microbiology**, v. 40, n. 6, p. 1994–2001, jun. 2002.

VAN LOO, I. H. M.; MOOI, F. R. Changes in the Dutch Bordetella pertussis population in the first 20 years after the introduction of whole-cell vaccines. **Microbiology (Reading, England)**, v. 148, n. Pt 7, p. 2011–8, jul. 2002.

WANG, Z. et al. High prevalence of erythromycin-resistant Bordetella pertussis in Xi'an, China. **Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases**, 11 maio 2014.

WARFEL, J. M.; ZIMMERMAN, L. I.; MERKEL, T. J. Acellular pertussis vaccines protect against disease but fail to prevent infection and transmission in a nonhuman primate model. **Proceedings of the National Academy of**

**Sciences of the United States of America**, v. 111, n. 2, p. 787–92, 14 jan. 2014.

WILLEMS, R. J. et al. Isolation of a putative fimbrial adhesin from *Bordetella pertussis* and the identification of its gene. **Molecular microbiology**, v. 9, n. 3, p. 623–34, ago. 1993.

WILLIAMSON, P.; MATTHEWS, R. Epitope mapping the Fim2 and Fim3 proteins of *Bordetella pertussis* with sera from patients infected with or vaccinated against whooping cough. **FEMS immunology and medical microbiology**, v. 13, n. 2, p. 169–78, fev. 1996.

ZHANG, S. et al. Complete genome sequence of *Bordetella pertussis* CS, a Chinese pertussis vaccine strain. **Journal of bacteriology**, v. 193, n. 15, p. 4017–8, 1 ago. 2011.

ZHANG, Z. et al. A greedy algorithm for aligning DNA sequences. **Journal of computational biology : a journal of computational molecular cell biology**, v. 7, n. 1-2, p. 203–14, 2000.

# **Anexos**

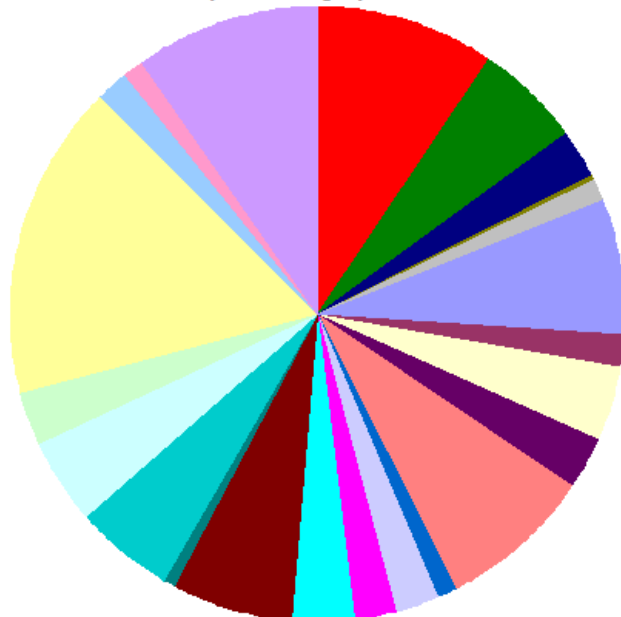
## **8. Anexos**

Esta sessão contém imagens com a caracterização funcional, baseada em subsistemas gerada pelo RAST, dos 4 genomas sequenciados neste estudo.

**Subsystem Coverage**



**Subsystem Category Distribution**



**Subsystem Feature Counts**

- ☐ Cofactors, Vitamins, Prosthetic Groups, Pigments (259)
- ☐ Cell Wall and Capsule (151)
- ☐ Virulence, Disease and Defense (67)
- ☐ Potassium metabolism (13)
- ☐ Photosynthesis (0)
- ☐ Miscellaneous (24)
- ☐ Phages, Prophages, Transposable elements, Plasmids (2)
- ☐ Membrane Transport (198)
- ☐ Iron acquisition and metabolism (47)
- ☐ RNA Metabolism (105)
- ☐ Nucleosides and Nucleotides (77)
- ☐ Protein Metabolism (217)
- ☐ Cell Division and Cell Cycle (30)
- ☐ Motility and Chemotaxis (61)
- ☐ Regulation and Cell signaling (59)
- ☐ Secondary Metabolism (4)
- ☐ DNA Metabolism (94)
- ☐ Regulons (7)
- ☐ Fatty Acids, Lipids, and Isoprenoids (173)
- ☐ Nitrogen Metabolism (15)
- ☐ Dormancy and Sporulation (3)
- ☐ Respiration (141)
- ☐ Stress Response (119)
- ☐ Metabolism of Aromatic Compounds (81)
- ☐ Amino Acids and Derivatives (448)
- ☐ Sulfur Metabolism (47)
- ☐ Phosphorus Metabolism (28)
- ☐ Carbohydrates (248)

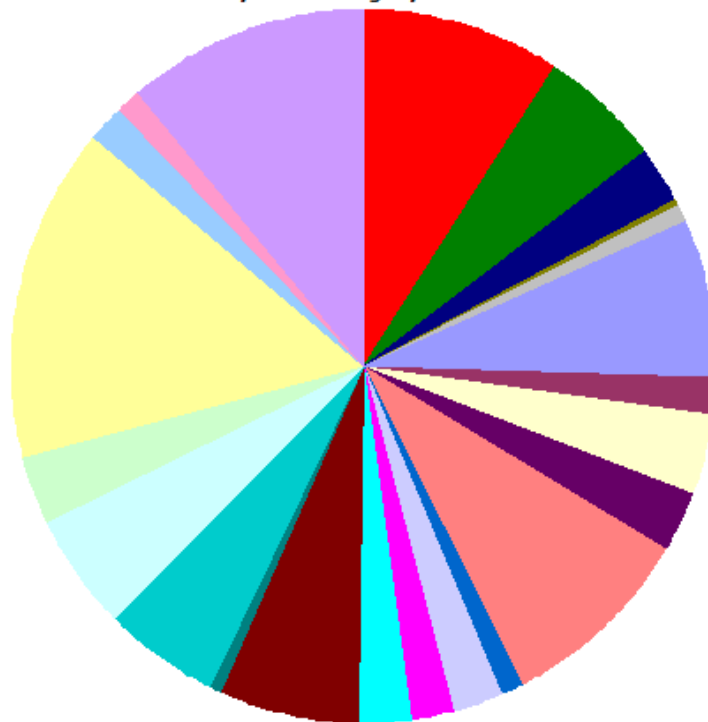
**Anexo I.I. Classificação funcional do subsistemas do RAST do genoma Bp168**



### Subsystem Coverage



### Subsystem Category Distribution



### Subsystem Feature Counts

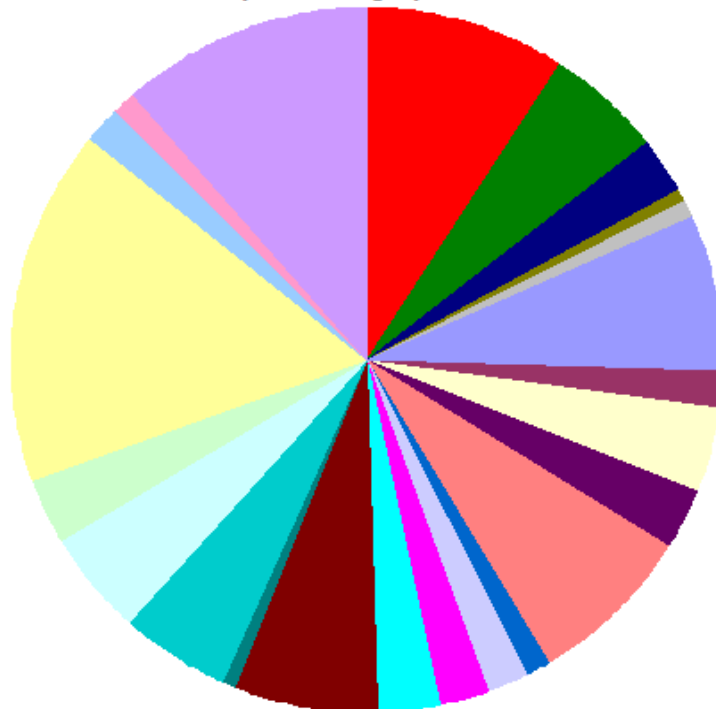
- ⊕ Cofactors, Vitamins, Prosthetic Groups, Pigments (258)
- ⊕ Cell Wall and Capsule (150)
- ⊕ Virulence, Disease and Defense (66)
- ⊕ Potassium metabolism (12)
- ⊕ Photosynthesis (0)
- ⊕ Miscellaneous (24)
- ⊕ Phages, Prophages, Transposable elements, Plasmids (2)
- ⊕ Membrane Transport (191)
- ⊕ Iron acquisition and metabolism (46)
- ⊕ RNA Metabolism (105)
- ⊕ Nucleosides and Nucleotides (76)
- ⊕ Protein Metabolism (241)
- ⊕ Cell Division and Cell Cycle (29)
- ⊕ Motility and Chemotaxis (60)
- ⊕ Regulation and Cell signaling (56)
- ⊕ Secondary Metabolism (4)
- ⊕ DNA Metabolism (70)
- ⊕ Regulons (7)
- ⊕ Fatty Acids, Lipids, and Isoprenoids (176)
- ⊕ Nitrogen Metabolism (15)
- ⊕ Dormancy and Sporulation (3)
- ⊕ Respiration (141)
- ⊕ Stress Response (148)
- ⊕ Metabolism of Aromatic Compounds (79)
- ⊕ Amino Acids and Derivatives (423)
- ⊕ Sulfur Metabolism (44)
- ⊕ Phosphorus Metabolism (27)
- ⊕ Carbohydrates (288)

Anexo I.II. Classificação funcional do subsistemas do RAST do genoma Bp100

### Subsystem Coverage



### Subsystem Category Distribution



### Subsystem Feature Counts

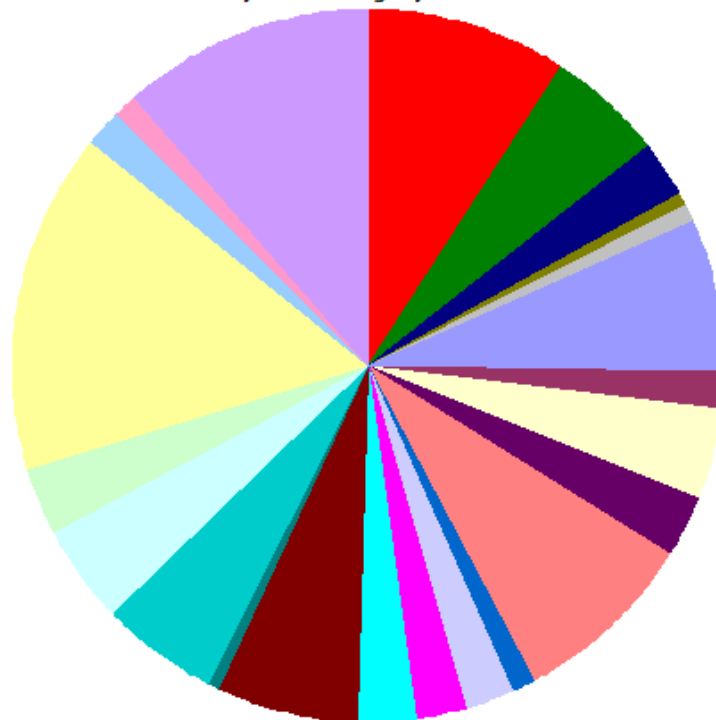
- ⊕ Cofactors, Vitamins, Prosthetic Groups, Pigments (256)
- ⊕ Cell Wall and Capsule (151)
- ⊕ Virulence, Disease and Defense (66)
- ⊕ Potassium metabolism (12)
- ⊕ Photosynthesis (0)
- ⊕ Miscellaneous (24)
- ⊕ Phages, Prophages, Transposable elements, Plasmids (2)
- ⊕ Membrane Transport (198)
- ⊕ Iron acquisition and metabolism (48)
- ⊕ RNA Metabolism (107)
- ⊕ Nucleosides and Nucleotides (77)
- ⊕ Protein Metabolism (208)
- ⊕ Cell Division and Cell Cycle (29)
- ⊕ Motility and Chemotaxis (61)
- ⊕ Regulation and Cell signaling (58)
- ⊕ Secondary Metabolism (4)
- ⊕ DNA Metabolism (78)
- ⊕ Regulons (7)
- ⊕ Fatty Acids, Lipids, and Isoprenoids (182)
- ⊕ Nitrogen Metabolism (15)
- ⊕ Dormancy and Sporulation (3)
- ⊕ Respiration (139)
- ⊕ Stress Response (133)
- ⊕ Metabolism of Aromatic Compounds (80)
- ⊕ Amino Acids and Derivatives (457)
- ⊕ Sulfur Metabolism (46)
- ⊕ Phosphorus Metabolism (27)
- ⊕ Carbohydrates (298)

Anexo I.III. Classificação funcional dos subsistemas do RAST do genoma Bp1913

### Subsystem Coverage



### Subsystem Category Distribution



### Subsystem Feature Counts

- ⊕ Cofactors, Vitamins, Prosthetic Groups, Pigments (249)
- ⊕ Cell Wall and Capsule (138)
- ⊕ Virulence, Disease and Defense (66)
- ⊕ Potassium metabolism (13)
- ⊕ Photosynthesis (0)
- ⊕ Miscellaneous (24)
- ⊕ Phages, Prophages, Transposable elements, Plasmids (2)
- ⊕ Membrane Transport (186)
- ⊕ Iron acquisition and metabolism (47)
- ⊕ RNA Metabolism (105)
- ⊕ Nucleosides and Nucleotides (76)
- ⊕ Protein Metabolism (224)
- ⊕ Cell Division and Cell Cycle (30)
- ⊕ Motility and Chemotaxis (61)
- ⊕ Regulation and Cell signaling (58)
- ⊕ Secondary Metabolism (4)
- ⊕ DNA Metabolism (71)
- ⊕ Regulons (7)
- ⊕ Fatty Acids, Lipids, and Isoprenoids (174)
- ⊕ Nitrogen Metabolism (13)
- ⊕ Dormancy and Sporulation (3)
- ⊕ Respiration (139)
- ⊕ Stress Response (122)
- ⊕ Metabolism of Aromatic Compounds (80)
- ⊕ Amino Acids and Derivatives (413)
- ⊕ Sulfur Metabolism (44)
- ⊕ Phosphorus Metabolism (27)
- ⊕ Carbohydrates (285)

Anexo I.IV. Classificação funcional do subsistemas do RAST do genoma Bp2041