

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**MÉTODO, SOFTWARE E BANCO DE DADOS PARA SOROTIPAGEM
MOLECULAR DE *Streptococcus pneumoniae* VISANDO O MONITORAMENTO DA
EFICÁCIA DO PROGRAMA DE VACINAÇÃO NO BRASIL**

por

Dhian Renato Almeida Camargo

Belo Horizonte

Fevereiro/2014

DISSERTAÇÃO MBCM - CPqRR

D.R.A. CAMARGO

2014

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**MÉTODO, SOFTWARE E BANCO DE DADOS PARA SOROTIPAGEM
MOLECULAR DE *Streptococcus pneumoniae* VISANDO O MONITORAMENTO DA
EFICÁCIA DO PROGRAMA DE VACINAÇÃO NO BRASIL**

por

Dhian Renato Almeida Camargo

**Dissertação apresentada com vistas à obtenção
do Título de Mestre em Ciências na área de
concentração Biologia Celular e Molecular.**

Orientação: Dr. Roney Santos Coimbra

Belo Horizonte
Fevereiro/2014

Catálogo-na-fonte
Rede de Bibliotecas da FIOCRUZ
Biblioteca do CPqRR
Segemar Oliveira Magalhães CRB/61975

C172m Camargo, Dhian Renato Almeida.
2014

Método, software e banco de dados para sorotipagem molecular de *Streptococcus pneumoniae* visando o monitoramento da eficácia do programa de vacinação no Brasil / Dhian Renato Almeida Camargo. – Belo Horizonte, 2014.

XXI, 145 f.: il.; 210 x 297mm.

Bibliografia: f.: 158 - 166

Dissertação (Mestrado) – Dissertação para obtenção do título de Mestre em Ciências pelo Programa de Pós - Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular.

1. Pneumonia Pneumocócica/imunologia 2. *Streptococcus pneumoniae*/enzimologia 3. Sorotipagem/método I. Título. II. Coimbra, Roney Santos (Orientação).

CDD – 22. ed. – 616.241

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**MÉTODO, SOFTWARE E BANCO DE DADOS PARA SOROTIPAGEM
MOLECULAR DE *Streptococcus pneumoniae* VISANDO O MONITORAMENTO DA
EFICÁCIA DO PROGRAMA DE VACINAÇÃO NO BRASIL**

por

Dhian Renato Almeida Camargo

Foi avaliada pela banca examinadora composta pelos seguintes membros:

Prof. Dr. Roney Santos Coimbra (Presidente)

Prof. Dr. Rodrigo Pedro Pinto Soares

Prof. Dr. Luiz Guilherme Dias Heneine

Suplente: Dr. Jerônimo Conceição Ruiz

Dissertação defendida e aprovada em: 27/02/2014.

COLABORADORES

Centro de Pesquisas René Rachou – Belo Horizonte

Dr. Guilherme Correa de Oliveira

Dra. Ângela Cristina Volpini

Dr. Fabiano Sviatopolk Mirsky Pais

Fundação Ezequiel Dias

Dra. Marluce Aparecida Assunção Oliveira

SUPORTE FINANCEIRO:

Centro de Pesquisas René Rachou/FIOCRUZ;

Fundação Ezequiel Dias - FUNED;

Fundação de Amparo à Pesquisa do Estado de Minas Gerais (Número do processo FAPEMIG - n° PPM-00614-11);

Conselho Nacional de Desenvolvimento Científico e Tecnológico (Número do processo CNPq – n° 479543/2012-7).

DEDICATÓRIA

Ofereço este trabalho àqueles que compartilham da minha busca a anos, Elio e Marta, que sempre me apoiaram em todas as minhas decisões e a todos que contribuíram para sua realização. Feliz são os filhos que possuem pais como vocês.

AGRADECIMENTOS

À Deus pela sua infinita bondade e amor.

Em especial, quero agradecer aos meus familiares, aos meus pais Elio Renato de Almeida e Maria Marta Camargo Almeida, ao meu irmão Renan Henning Camargo Almeida e minha esposa Marília Pereira da Costa pelo incentivo e carinho.

Ao meu orientador, Dr. Roney Santos Coimbra, muito obrigado pelo apoio, incentivo, ensinamentos, dedicação e atenção mais que especial nas revisões e sugestões para a realização deste trabalho.

À Mestre Marluce Aparecida Assunção Oliveira, que tornou possível a realização deste trabalho e aos colegas do Serviço de Doenças Bacterianas e Fúngicas, em especial Vanda Lúcia Fagundes de Souza, Carmem Dolores Faria, Michelle Lara Samuel e Max Assunção Correia.

Aos colegas do LPCM – Laboratório de Parasitologia Celular e Molecular e do CEBio – Centro de Excelência em Bioinformática, em especial a Dra Ângela Volpini, Dr. Guilherme Corrêa Oliveira, Dr. Fabiano Sviatopolk-Mirsky Pais, Francislton Silva de Oliveira, Fausto Gonçalves dos Santos, Barbara Rezende Pereira da Mata e demais colaboradores.

Aos colegas de classe, pela amizade e rica troca de experiências.

A todos que, de alguma forma contribuíram para esta construção.

À Fundação Oswaldo Cruz, ao Centro de Pesquisas René Rachou e ao Programa de Pós-graduação em Ciências da Saúde pela oportunidade ofertada.

À Fundação Ezequiel Dias por possibilitar a realização deste trabalho e pelo aprendizado.

A Fundação de Amparo a Pesquisa do Estado de Minas Gerais (FAPEMIG) pela bolsa concedida através do Programa de Capacitação de Recursos Humanos do Estado de Minas Gerais (PCRH) e pelo financiamento do projeto de pesquisa (Número do processo FAPEMIG - n° PPM-00614-11).

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo financiamento do projeto de pesquisa (Número do processo CNPq – n° 479543/2012-7).

Sumário

Lista de figuras	XIII
Lista de tabelas	XV
Lista de Abreviaturas e Símbolos	XVI
Resumo	XX
Abstract	XXI
1 INTRODUÇÃO	22
1.1 <i>Streptococcus pneumoniae</i>	23
1.2 Infecções por <i>Streptococcus pneumoniae</i>	24
1.3 Vacinas contra <i>S. pneumoniae</i>	26
1.4 <i>S. pneumoniae</i> e o <i>locus cps</i>	31
1.5 Sorotipagem de <i>S. pneumoniae</i>	36
1.6 Sorotipagem molecular	38
1.7 Justificativa	43
2 OBJETIVO	45
2.1 Objetivo Geral.....	46
2.2 Objetivos específicos	46
3 MATERIAL E MÉTODOS	47
3.1 Local do estudo	48
3.2 Sequências de DNA da região <i>cps</i> de <i>S. pneumoniae</i>	48
3.3 Procedimentos experimentais	48
3.3.1 Análises <i>in silico</i>	48
3.3.2 Padronização do método experimental	50
3.4 Análise estatística	54
3.5 Criação do banco de dados com perfis <i>cps</i> -RFLP	54
3.6 Requisitos legais	55
4 RESULTADOS.....	56
4.1 Resultados das análises <i>in silico</i>	57
4.2 Validação da enzima de restrição selecionada frente à endonuclease <i>HinfI</i>	65
4.3 Validação experimental	68

4.3.1	Amplificação e restrição do cluster <i>cps</i>	68
4.3.2	A separação dos fragmentos e a estimativa de seus pesos moleculares usando o sistema Agilent 2100 Bioanalyzer	78
4.4	Trabalhos apresentados em congressos científicos.....	79
5	DISCUSSÃO	80
6	CONCLUSÕES	88
7	PERSPECTIVAS.....	90
8	ANEXOS	92
	<i>ANEXO I – Relação das cepas utilizadas na etapa de validação experimental.</i>	<i>93</i>
	<i>ANEXO II – Banco de dados gerado na etapa de análise in silico pela enzima XhoII. ...</i>	<i>97</i>
	<i>ANEXO III – Banco de dados gerado na etapa de análise experimental pela enzima XhoII.....</i>	<i>100</i>
	<i>ANEXO IV – Tamanho em pares de bases dos produtos de amplificação da região cps na análise experimental.....</i>	<i>102</i>
	<i>ANEXO V – Resultados das análises dos perfis experimentais cps-RFLP testados contra o banco de dados carregado no software MST.....</i>	<i>103</i>
	<i>ANEXO VI – Banco de dados gerado na etapa de análise experimental no sistema automatizado de eletroforese (BioAnalyzer).....</i>	<i>105</i>
	<i>ANEXO VII – Requisito legal.....</i>	<i>108</i>
	<i>ANEXO VIII – Resumo e o certificado de apresentação do trabalho acadêmico no XXI Congresso Latinoamericano de Microbiologia (XXI ALAM).....</i>	<i>109</i>
	<i>ANEXO IX – Resumo e o certificado de apresentação do trabalho acadêmico no X-meeting 2013.</i>	<i>111</i>
	<i>ANEXO X – Script em linguagem Perl utilizado para download das sequências de DNA da região cps</i>	<i>114</i>
	<i>ANEXO XI – Script em linguagem Perl utilizado para identificar as endonucleases com sítios de restrição em todas as sequências de DNA da região cps</i>	<i>119</i>
	<i>ANEXO XII - Script em linguagem Perl utilizado para capturar os identificadores e ordenar os fragmentos de digestão in silico</i>	<i>123</i>
	<i>ANEXO XII - Script em linguagem Perl utilizado para gerar as matrizes de distância entre os perfis de restrição.....</i>	<i>137</i>
	<i>ANEXO XIV - Script em linguagem Perl utilizado para gerar as figuras do tamanho dos produtos de amplificação da região cps</i>	<i>146</i>

<i>ANEXO XV - Script em linguagem Perl utilizado para gerar as figuras com os perfis de restrição da região cps.....</i>	<i>152</i>
9 REFERÊNCIAS BIBLIOGRÁFICAS	158

Lista de figuras

Figura 1: Representação da via <i>dependente</i> W_{zx}/W_{zy} para biossíntese do CPS 9A.	33
Figura 2: Representação esquemática de alguns clusters CPS e das respectivas unidades repetidas.....	35
Figura 3: Fluxograma mostrando a estratégia para sorotipagem clássica de pneumococos (<i>chessboard</i>).	37
Figura 4: <i>Molecular Serotyping Tool</i> for <i>E. coli/Shigella spp</i> acessível nos servidores do CEBio (http://www.cebio.org/mst).	41
Figura 5: Resultado esquemático após análise do perfil experimental pelo Software MST...42	
Figura 6: Representação esquemática do dendograma de distância entre os perfis de restrição gerados pela enzima BslFI.....	58
Figura 7: Representação esquemática do dendograma de distância entre os perfis de restrição gerados pela enzima Eco57MI.	59
Figura 8: Representação esquemática do dendograma de distância entre os perfis de restrição gerados pela enzima HindII.....	60
Figura 9: Representação esquemática do dendograma de distância entre os perfis de restrição gerados pela enzima StyI.	61
Figura 10: Representação esquemática do dendograma das distâncias entre os perfis de restrição gerados pela enzima XhoII.	62
Figura 11: Resultado da análise para escolha do score MST mais distinguível.....	64
Figura 12: Resultado das análises para escolher a faixa de peso molecular mais discriminante.	64
Figura 13: A substituição da endonuclease HinfI por XhoII aumentou a especificidade do <i>cps</i> -RFLP.....	66
Figura 14: Representação esquemática do dendograma de distância entre os perfis de restrição gerados pela enzima HinfI.	67
Figura 15: Produto da amplificação do cluster <i>cps</i> dos sorotipos 19F, 18C, 19A, 6B, 18B, 14.	68
Figura 16: Perfis de restrição com a enzima XhoII <i>cps</i> -RFLP dos sorotipos 29, 9N, 19F, 6C, 19A, 29, 6A, 19F, 18B, 6B.....	69
Figura 17: Produto da amplificação do cluster <i>cps</i> dos sorotipos 24F, 12F, 35B, 5, 10A, 9V, 9N e 7F.	70
Figura 18: Perfis de restrição com a enzima XhoII <i>cps</i> -RFLP dos sorotipos 24F, 12F, 35B, 5, 18C, 10A, 9V e 7F.....	71

Figura 19: Produto da amplificação do cluster <i>cps</i> dos sorotipos 6C, 35B, 9N, 3, 6C, 10A, 6A/C, 9N e 6A/C.	72
Figura 20: Perfis de restrição com a enzima XhoII <i>cps</i> -RFLP dos sorotipos 6C, 35B, 9N, 10A, 6A/C, 9N e 6A/C.	73
Figura 21: O perfil <i>cps</i> -RFLP experimental obtido demonstra estreita similaridade com aquele previsto <i>in silico</i>	74
Figura 22: Representação esquemática do dendograma de distância entre os perfis de restrição experimentais gerados pela enzima XhoII.	75
Figura 23: Representação esquemática dos dois pares de perfis experimentais indistinguíveis e do par <i>in silico</i> /experimental indistinguível.	77
Figura 24: Representação esquemática do eletroferograma do perfil de restrição experimental gerado pela enzima XhoII da amostra 317/08 (sorotipo 34)	79

Lista de tabelas

Tabela 1: Classificação de <i>Streptococcus pneumoniae</i> em sorogrupos e sorotipos.	30
Tabela 2: Resultado das análises estatísticas realizadas para seleção da enzima de restrição mais discriminante.	57
Tabela 3: Resultado da análise estatística para comparar as enzimas XhoII e HinfI.	65
Tabela 4: Resultado das análises dos perfis experimentais contra o banco de dados virtual gerado na etapa de análise <i>in silico</i> utilizando o software MST.	76
Tabela 5: Resultados da comparação entre os métodos <i>cps</i> -RFLP, PCR-RFLP e sorotipagem clássica.	85

Lista de Abreviaturas e Símbolos

A - Adenina

ABNT - Associação Brasileira de Normas Técnicas

ALAM - Associação Latino Americana de Microbiologia

BHI - *Brain Heart Infusion*

C - Citosina

CDC - *Centers for Disease Control and Prevention*

CEBio - Centro de Excelência em Bioinformática

CIE - Reação de contraímuno eletroforese

cm - Centímetro

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

COA - Teste de coagulação

CPqRR - Centro de Pesquisas René Rachou

CPS - Cápsula de polissacarídeos

CRIE - Centros de Referência de Imunobiológicos Especiais

DGP - Diplococos Gram positivos

DMSO - Dimetilsulfóxido

DNA - Ácido desoxirribonucleico

dNTPs - *Deoxyribonucleotide Triphosphates*

DPI - Doença pneumocócica invasiva

EDTA - Ácido etilenodiaminatetracético

EMBOSS - *European Molecular Biology Open Software Suite*

EVTECIAS - Estudo de Viabilidade Técnica, Econômica, Comercial e de Impactos

Ambientais e Sociais

FAPEMIG - Fundação de Amparo à Pesquisa do estado de Minas Gerais

FIOCRUZ - Fundação Osvaldo Cruz

Funed - Fundação Ezequiel Dias

G - Guanina

GSK - GlaxoSmithKline

h - Horas

HCl - Ácido clorídrico

HGs - Grupos homólogos

IAL - Instituto Adolfo Lutz

JPEG - *Joint Photographic Experts Group*

LA - Teste de aglutinação do látex

LACEN - Laboratório Central de Saúde Pública

LACEN/MG - Laboratório Central de Saúde Pública do Estado de Minas Gerais

LACENS - Laboratórios Centrais de Saúde Pública

LPCM - Laboratório de Parasitologia Celular e Moleular

mg - Miligrama

MG - Estado de Minas Gerais

Min - Minuto

mL - Mililitro

MLST - *Multilocus Sequence Typing*

MS - Ministério da Saúde

MST - *Molecular Serotyping Tool*

NaCl - Cloreto de sódio

NT - Não-tipáveis

NTQI - Núcleo de Tecnologia da Qualidade da Inovação

OMS - Organização Mundial da Saúde

PA - Pureza analítica

pb - Pares de bases

PCR - Reação em cadeia da polimerase

PERL - *Practical Extraction and Report Language*

PFGE - *Pulsed-field Gel Electrophoresis*

PHYLIP - *PHYLogeny Inference Package*

PII - Programa de Incentivo à Inovação

PNI - Programa Nacional de Imunização

RFLP - *Restriction Fragment Length Polymorphism*

RNA - Ácido ribonucleico

rpm - Rotações por minuto

SINAN - Sistema de Informação de Agravos de Notificação

SNC - Sistema nervoso central

SP - São Paulo

Spn - *Streptococcus pneumoniae*

SSI - Statens Serum Institut

SUS - Sistema Único de Saúde

T - Timina

TAE - Tris-acetato-EDTA

TBE - Tris-borato-EDTA

TE - Tris-EDTA

TIFF - *Tagged Image File Format*

U - Unidade

V - Volts

VPC7 - Vacina anti-pneumocócica conjugada 7-valente

VPC10 - Vacina anti-pneumocócica conjugada 10-valente

VPC13 - Vacina anti-pneumocócica conjugada 13-valente

VPP23V - Vacina anti-pneumocócica polissacarídea 23-valente

WHO - *World Health Organization*

% - Percentagem

°C - Grau Celsius

μL - Microlitro

μm - Micrômetro

Resumo

Noventa e dois sorotipos de *Streptococcus pneumoniae* já foram descritos, mas a vacina pneumocócica conjugada (PCV10) introduzida no calendário básico de vacinação do Brasil, em 2010, cobre somente os mais prevalentes no país. A substituição dos sorotipos vacinais após imunização em massa é uma grande preocupação e monitorar esse fenômeno requer métodos de sorotipagem eficientes e acessíveis. A Sorotipagem clássica de pneumococos baseada em antissoros produzidos em animais é trabalhosa, restrita a poucos laboratórios de referência, e não pode tipar isolados não capsulados. Alternativamente, os métodos de sorotipagem molecular avaliam os polimorfismos dos genes do cluster *cps*, que codificam enzimas chave para a síntese do CPS em *Streptococcus pneumoniae*. Em uma abordagem apropriada, *cps*-RFLP, os loci *cps* amplificados por PCR são digeridos com uma endonuclease gerando perfis únicos à eletroforese em gel de agarose, permitindo assim a identificação do sorotipo. Neste trabalho, nós combinamos abordagens *in silico* e *in vitro* para demonstrar que XhoII é a endonuclease mais discriminante para o método *cps*-RFLP, e para construir um banco de dados de perfis sorotipo-específico que acomodou a diversidade genética do locus *cps* de 91 conhecidos sorotipos de pneumococos. O banco de dados de perfis *cps*-RFLP foi integrado ao *Molecular Serotyping Tool* (MST), software anteriormente publicado baseado em *web-based* para sorotipagem molecular. Usando XhoII, o método *cps*-RFLP obteve especificidade de 84,6% para sorotipagem e 100 % para sorogrupagem de pneumococos. Esta nova ferramenta pode representar uma colaboração relevante para vigilância epidemiológica em tempo real da diversidade de pneumococos em resposta a programas de imunização em massa.

Abstract

Ninety-two *Streptococcus pneumoniae* serotypes have been described, but the pneumococcal conjugate vaccine (PCV10) introduced in the *Brazilian basic vaccination schedule*, in 2010, covers only the most prevalent in the country. Pneumococcal serotype-shifting after massive immunization is a major concern and monitoring this phenomenon requires efficient and accessible serotyping methods. Classical pneumococcal serotyping based on antisera produced in animals is laborious, restricted to a few reference laboratories, and cannot type non-capsulated isolates. Alternatively, molecular serotyping methods assess the polymorphisms in the *cps* gene cluster, which encodes key enzymes for CPS synthesis in *Streptococcus pneumoniae*. In one such approach, *cps*-RFLP, the PCR amplified *cps* loci are digested with an endonuclease, generating unique fingerprints on agarose gel electrophoresis, therefore allowing serotype identification. In this work, we combined *in silico* and *in vitro* approaches to demonstrate that XhoII is the most discriminating endonuclease for *cps*-RFLP, and to build a database of serotype-specific fingerprints that accommodates the genetic diversity within the *cps* locus of 91 known pneumococci serotypes. The database of *cps*-RFLP fingerprints was integrated to Molecular Serotyping Tool (MST), our previously published web-based software for molecular serotyping. Using XhoII, the *cps*-RFLP method achieved 84.6% specificity for serotyping and 100% for serogrouping of pneumococci. This new tool may represent a relevant aid to real time epidemiological surveillance of pneumococci diversity in response to mass immunization programs.

1.1 *Streptococcus pneumoniae*

A classificação taxonômica do *Streptococcus pneumoniae*, a partir do táxon mais amplo é: Domínio *Bacteria*, Filo *Firmicutes*, Classe *Bacilli*, Ordem *Bacillales*, Família *Streptococcaceae*, Gênero *Streptococcus*, Espécie *Streptococcus pneumoniae*. *S. pneumoniae* é um membro do grupo *Streptococcus mitis-Streptococcus oralis* (o grupo Smit) do *Streptococcus* grupo viridans, que inclui *S. mitis*, *S. oralis*, *S. cristatus*, *Streptococcus infantis*, e *Streptococcus peroris*. *Streptococcus pneumoniae* (pneumococo) é um coco Gram positivo encapsulado, medindo entre 0,5 e 1,25 µm, que ocorre tipicamente aos pares, ocasionalmente isolados ou em cadeias curtas. A extremidade de cada célula no par tende a ser ponteadada ou lanceolada. Sucessivos cultivos no laboratório promovem a formação de cadeias. Uma reação Gram positiva em células jovens pode ser perdida em culturas velhas que adquirem o aspecto de Gram negativas. No isolamento primário, o pneumococo apresenta-se fortemente encapsulado com polissacarídeos e as colônias mucóides resultam do aumento da síntese do polissacarídeo capsular. Colônias rugosas ocorrem raramente e este fenótipo se deve à interrupção da produção dos polissacarídeos da cápsula. As colônias de pneumococo geralmente apresentam-se pequenas, brilhantes, achatadas e umbilicadas, produzem alfa-hemólise em ágar sangue de carneiro à 5% após incubação por 18-24 horas em atmosfera de 5-10% de CO², e não produzem catalase nem oxidase (Guerrero, 2001).

A diferenciação de pneumococo dos *Streptococcus* do grupo viridans, por convenção, é realizada através da demonstração *in vitro* da susceptibilidade à optoquina e/ou solubilidade em bile (desoxicolato de sódio). A identificação por meio destes dois testes é satisfatória quando isolados de sítios estéreis são testados. O teste de susceptibilidade à optoquina é o teste mais comum utilizado nos laboratórios de microbiologia clínica para triar colônias alfa-hemolíticas para identificação fenotípica do pneumococo. No entanto, a resistência à optoquina tem sido relatada em *Streptococcus pneumoniae* e laboratórios que utilizam apenas este teste podem ter problemas na identificação deste microrganismo, principalmente quando associado a colônias com morfologia atípica. Cepas insolúveis em bile foram relatadas e também cepas atípicas resistentes à optoquina e insolúveis em bile. Apesar destes relatos, o teste de solubilidade em bile apresenta sensibilidade superior a 98% e especificidade de 100% (Arbique *et al.*, 2004a).

Sistemas automatizados para identificação de pneumococos estão disponíveis no mercado. Os sistemas automatizados para identificação e teste de sensibilidade VITEK 2 (Biomèrieux, Marcy l'Etoile, França), BD Phoenix (Bacton Dickinson Labware, Franklin

Lakes, NJ) e MicroScan (Siemens, Erlange, Alemanha) são os mais utilizados (Ligozzi *et al.*, 2002; Richter *et al.*, 2007; Mittman *et al.*, 2009).

Testes rápidos de aglutinação do látex para detecção sorológica de pneumococo também estão disponíveis. Esses métodos dependem da detecção do complexo antígeno anticorpo resultante da reação entre o antígeno de superfície do pneumococo com o anticorpo tipo específico. Todos estes testes utilizam o *Omnim Serum* do *Statens Serum Institut*, reagente que contém todos os anticorpos capsulares para todos os sorotipos de pneumococo conhecidos. Pneumococos não capsulados não são detectados por estes métodos (Slotved *et al.*, 2004).

Técnicas moleculares como PCR para detecção de genes de virulência do pneumococo (autolisina [*lytA*] e pneumolisina [*ply*]) e dos genes da proteína ligante de penicilina (PBPs), assim como os testes de hibridização de sondas em regiões específicas do gene 16S rRNA (16S rDNA) têm sido utilizados para identificação do pneumococo (Arbique *et al.*, 2004a; Carvalho *et al.*, 2007). O sequenciamento do gene 16S rRNA tem uso limitado para identificação de *S. pneumoniae* devido à alta similaridade genética entre as espécies do grupo Smit (Arbique *et al.*, 2004a). Identificação de *S. pneumoniae* com *Matrix-assisted laser desorption/ionisation-time of flight* (MALDI-TOF) *mass spectrometry* vem sendo proposta para utilização na rotina de laboratórios de diagnóstico microbiológico. Este método oferece um modo rápido e simples de identificar o pneumococo, mas a sua principal desvantagem ainda é o custo elevado do equipamento (Branda *et al.*, 2013; Dubois *et al.*, 2013).

1.2 Infecções por *Streptococcus pneumoniae*

Streptococcus pneumoniae é um importante patógeno humano, sendo considerado o principal agente bacteriano causador de pneumonia e meningite bacteriana. Este patógeno ganha acesso ao hospedeiro colonizando a mucosa da nasofaringe humana. Com a maioria dos indivíduos colonizados, a bactéria mantém uma relação comensal, sendo esses indivíduos apontados como portadores assintomáticos. Os portadores assintomáticos representam um importante reservatório para a transmissão deste patógeno na comunidade. Em média, a taxa de colonização de indivíduos assintomáticos é de 30-50% em crianças menores de cinco anos e 10-20% em adultos (Ferreira *et al.*, 2001).

A infecção pneumocócica é uma das maiores causas de morbidade e mortalidade em todo o mundo, podendo se apresentar como doenças pneumocócicas não invasivas (sinusite, otite média aguda, conjuntivite, bronquite e pneumonia), ou graves doenças invasivas (pneumonias bacterêmicas, meningite, sépsis e artrite). Nas doenças pneumocócicas invasivas

(DPI), o patógeno invade regiões do hospedeiro que deveriam ser estéreis. A transmissão dos pneumococos se dá, geralmente, por contato direto entre indivíduos e é favorecida pela aglomeração de pessoas. Infecções respiratórias virais e clima frio podem facilitar a disseminação desse microrganismo. As crianças e os idosos são particularmente suscetíveis à doença pneumocócica invasiva. Infecções não invasivas como otites, sinusites e conjuntivites, também têm grande importância em saúde pública. Entre 30% e 40% das otites médias agudas, 40% das sinusites agudas e 60% das pneumonias bacterianas adquiridas na comunidade são causadas pelos pneumococos (Franco-Alvarez De Luna *et al.*, 2005; Pneumococcal conjugate vaccine for childhood immunization--WHO position paper., 2007). No Brasil, entre 2004 e 2006, a doença pneumocócica foi responsável por 34.217 internações no Sistema Único de Saúde brasileiro (0,1% de todas as internações do setor público). A pneumonia pneumocócica foi responsável por 64,8% destas internações e o pneumococo foi responsável por 31,3% das internações de todos os casos de meningite bacteriana na qual o patógeno foi identificado (Novaes, Sartori e Soárez, 2011).

Em 2000, a OMS informou a ocorrência de aproximadamente 14,5 milhões de casos de doença pneumocócica no mundo e cerca de 826 mil mortes em crianças de um a 59 meses, das quais 90 mil em crianças HIV positivas e 735 mil em crianças HIV negativas. Das mortes em crianças HIV negativas, mais de 61% ocorreram em 10 países da Ásia e África (O'Brien *et al.*, 2009). Ainda segundo a OMS, em 2005, 1,6 milhões de mortes foram causadas pelo pneumococo. Esta estimativa incluiu as mortes de 700 mil a 1.000.000 de crianças menores de cinco anos. A maioria das crianças tinha idade inferior a dois anos e viviam em países pobres. Na Europa e nos EUA, *S. pneumoniae* é a causa mais comum de pneumonia bacteriana adquirida na comunidade em adultos. Nessas regiões, a incidência anual de doença pneumocócica invasiva varia de 10 a 100 casos por 100 000 habitantes. Mesmo em regiões economicamente mais desenvolvidas, a taxa de mortalidade para adultos com pneumonia pneumocócica é de 10% a 20%, embora possa ser superior a 50% nos grupos de alto risco (<http://www.who.int/ith/diseases/pneumococcal/en/>). A Organização Pan-americana de Saúde, após análise dos dados epidemiológicos publicados entre 1990 e 2006 referente à América Latina e Caribe, constatou a ocorrência de 1.261.348 casos de otite média aguda, 268.432 casos de pneumonia, 3.918 casos de meningite e 1.229 casos de septicemia, resultando em mais de 18.000 óbitos anualmente. Portanto, as doenças pneumocócicas constituem um problema grave de saúde pública (Constenla, 2007).

1.3 Vacinas contra *S. pneumoniae*

Até hoje, foram descritos 92 sorotipos de pneumococos imunologicamente distintos. A distribuição dos sorotipos pode variar de acordo a região geográfica, a época, a manifestação clínica e a idade do paciente. Assim, no Brasil, no período de 1977 a 2000, foram identificados os sorotipos mais frequentes associados às doenças invasivas, são eles: 1, 5, 6A, 6B, 9V, 14, 18C, 19F e 23F. Das cepas isoladas, 76,5% foram encontradas em crianças menores de cinco anos de idade com meningite (Brandileone *et al.*, 2003).

A Rede SIREVA (Sistema Regional de Vacinas) e depois, SIREVA II foi consolidada como uma referência indispensável no estudo, prevenção e controle das infecções causadas por *Streptococcus pneumoniae*, *Haemophilus influenzae* e *Neisseria meningitidis* na América Latina e Caribe (LAC). Uma das intervenções mais eficazes no arsenal da saúde pública, a vacinação, é um fator fundamental que contribui para a saúde, equidade e a segurança humana e que por muito tempo tem sido uma das áreas mais importantes de cooperação técnica da Organização Pan-Americana da Saúde (OPS).

O informe regional da rede SIREVA II, 2012, divulgou os dados por país e por grupos de idade sobre as características dos isolamentos de *Streptococcus pneumoniae*, *Haemophilus influenzae* e *Neisseria meningitidis* de pacientes com doenças invasivas. Dos 1059 isolamentos de *S. pneumoniae* no Brasil, em 2012, 59,5% foram em pacientes do sexo masculino, 35,2% das cepas foram isoladas de pacientes com 50 anos ou mais e em 18,6% dos casos os pacientes tinham menos de cinco anos. Os casos com diagnóstico clínico de meningite responderam por 44,8% dos isolados e, por consequência, os isolamentos invasivos foram de 53,4% do material clínico sangue e de 38,6% de líquido. Em 197 cepas de pacientes menores de cinco anos os sorotipos 14 e 3 foram o mais prevalentes, cada um representou 14,2% dos isolamentos, seguidos pelos sorotipos 19A (9,6%), 6B (6,1%), 23F (5,6%), 6A (5,1%), 6C (4,1%), 11A (4,1%), 18C (3,6%), 22F (3,0%), 7F (2,5%), 12F (2,5%), 25A (2,5%), 15C (2,0%), 24F (2,0%), 9V (2,0%), 15A (1,5%), 20 (1,5%), 4 (1,0%), 19F (1,0%), 7C (1,0%), 8 (1,0%), 13 (1,0%), 16F (1,0%), 17F (1,0%), 18A (1,0%), 23A (1,0%), 23B (1,0%), 5 (0,5%), 9N (0,5%), 15B (0,5%), 18B (0,5%), 28A (0,5%), 29 (0,5%), 35B (0,5%) e em 832 cepas isoladas de pacientes com cinco anos ou mais o sorotipo 3 foi o mais prevalente com 10,0% dos isolamentos, seguidos pelos sorotipos 12F (8,9%), 23F (5,3%), 7F (4,8%), 14 (4,7%), 8 (4,4%), 4 (4,2%), 9N (3,7%), 19A (3,4%), 5 (3,0%), 6A (2,9%), 19F (2,9%), 6C (2,8%), 9V (2,6%), 18C (2,5%), 6B (2,4%), 10A (2,3%), 25A (2,3%), 11A (2,0%), 20 (2,0%), 22F (1,9%), 16F (1,8%), 13 (1,7%), 7C (1,4%), 18A (1,4%), 29 (1,4%), 15C (1,3%), 17F (1,2%), 35F (1,2%), 15A (1,1%), 18B (1,1%), 23B (1,0%), 15B (0,8%), 35B (0,8%), 1

(0,7%), 24F (0,7%), 31 (0,7%), 23A (0,6%), 11B (0,4%), 28A (0,4%), 35C (0,2%), 42 (0,2%), cepas não tipáveis (0,2%), 19B (0,1%), 21 (0,1%), 35A (0,1%). Dos pneumococos isolados de pacientes com de meningite, 25,7 % eram resistentes à penicilina e 3,7% eram resistentes a ceftriaxona, enquanto que 4,6% das cepas de casos de não meningite apresentaram somente resistência de nível intermediário à penicilina e 3,0% resistência de nível intermediário à ceftriaxona (OPS, 2013).

As vacinas de polissacarídeos capsulares anti-pneumocócicas estão licenciadas no mundo desde 1977, fazendo parte da lista de vacinas preconizadas pelo Ministério da Saúde: 1) a vacina pneumocócica polissacarídica 23-valente; 2) pneumocócica conjugada 7-valente; 3) pneumocócica conjugada 10-valente; 4) e a pneumocócica conjugada 13-valente. Os sorotipos cobertos por estas vacinas têm sido considerados de maior relevância epidemiológica para as doenças pneumocócicas no mundo. Alguns poucos estudos sugerem que cerca de 60 - 80% dos sorotipos de pneumococos responsáveis por doenças invasivas no Brasil são cobertos pela vacina polissacarídica 23-valente composta pelos sorotipos 1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F, 33F (Brandileone *et al.*, 1995; Mantese *et al.*, 2003), justificando sua incorporação pelo Programa Nacional de Imunizações, no ano de 1992, para grupos em condições clínicas especiais atendidos nos Centros de Referência de Imunobiológicos Especiais (CRIE). A Portaria nº 48, de 28 de julho de 2004, Ministério da Saúde, Brasil, instituiu as diretrizes gerais para o funcionamento dos CRIE, com a finalidade de facilitar o acesso da população, em especial dos portadores de imunodeficiência congênita ou adquirida e de outras condições especiais de morbidade ou exposição a situações de risco, aos imunobiológicos especiais para prevenção das doenças que são objeto do Programa Nacional de Imunizações – PNI. A partir de 1999, a vacina polissacarídica 23-valente passou a ser aplicada durante a Campanha Nacional de Vacinação do Idoso, visando pessoas de 60 anos ou mais que convivem em instituições de assistência ao idoso. A imunidade é adquirida cerca de 10 a 15 dias após a vacinação. No entanto, estudos demonstraram que crianças menores de dois anos, grupo de maior incidência da doença pneumocócica, podem não apresentar resposta adequada aos antígenos da vacina e os níveis de anticorpos estimulados podem não ser persistentes, pois as vacinas compostas por antígenos polissacarídeos não induzem resposta imunológica celular T (em geral, polissacarídeos são antígenos T-independentes que são capazes de estimular diretamente as células B a produzirem anticorpos sem a necessidade da célula T auxiliar) e, portanto, não geram memória imunológica (Mond, Lees e Snapper, 1995; Jones, 2005; WHO, 2007; CDC, 2008).

As vacinas conjugadas combinam quimicamente polissacarídeos capsulares e antígenos proteicos. A conjugação dos polissacarídeos a um carreador proteico gera um antígeno T-dependente que é capaz de induzir o nosso sistema imune a produzir uma resposta imune dependente de células T que resulta numa eficiente imunização de crianças menores de dois anos (Jefferies *et al.*, 2011). A proteína conjugada é processada em peptídeos pelas células apresentadoras de antígenos, que os apresenta aos linfócitos T CD4⁺ via MHC de classe II, o que resulta na proliferação e diferenciação de células T efetoras. Esses linfócitos T efetores são ativados e passam a produzir citocinas (IL-4, IL-5 e IL-6) que são responsáveis por estimular os linfócitos B a se diferenciarem em plasmócitos, produtores de imunoglobulina G, e a gerarem células B de memória (Pletz *et al.*, 2008; Avci *et al.*, 2011). A introdução das vacinas polissacarídicas conjugadas em programas de imunização em mais de 90 países tem reduzido consideravelmente os casos da doença em todo o mundo. Ao mesmo tempo, sorotipos não cobertos pelas vacinas conjugadas passaram a causar a maioria dos casos remanescentes de meningite pneumocócica nestes países, fenômeno denominado substituição de sorotipos (Reinert, Paradiso e Fritzell, 2010). A vacina conjugada anti-pneumocócica heptavalente (VPC-7) foi incorporada pelo Ministério da Saúde em 2001, também para grupos em condições clínicas especiais. Essa formulação cobre os sorotipos 4, 6B, 9V, 14, 18C, 19F e 23F.

A partir do ano de 2010, foi introduzida no calendário básico de vacinação a vacina pneumocócica conjugada decavalente (VPC10). Essa formulação da vacina contém todos os sorotipos presentes na VPC7 mais os sorotipos 1, 5, e 7F, conjugados a proteína D do hemófilus não tipável, e tem sido usada para imunizar crianças entre dois e 24 meses de idade protegendo-as contra doença invasiva e otite média aguda causada por *Streptococcus pneumoniae*. Em 2013, a Agência Nacional de Vigilância Sanitária (ANVISA) aprovou a indicação para adultos a partir de 50 anos da vacina pneumocócica conjugada 13-valente (VPC13), esta formulação contém todos o sorotipos da 10-valente, acrescida dos sorotipos 3, 6A e 19A. A VPC13 está disponível apenas na rede particular. A vacina conjugada 10-valente foi introduzida na rotina de vacinação do Brasil em 2010. A disponibilização da vacina VPC-10 se deve ao acordo de transferência tecnológica entre o laboratório GlaxoSmithKline – GSK e o Instituto de Tecnologia em Imunobiológicos Bio-Manguinhos/FIOCRUZ que passou a produzir esta vacina para todo o país. A cobertura vacinal, no ano de 2011, para menores de um ano foi de 81,5% (OPS, 2012).

Como resultado da inclusão da vacina conjugada decavalente na rotina de vacinação infantil no SUS, espera-se uma redução importante no número de casos de doenças

pneumocócicas nos próximos anos. No entanto, existe a possibilidade de que, após a imunização em massa no país, novos sorotipos possam preencher o nicho anteriormente ocupado pelos sorotipos eliminados. Este fenômeno, denominado substituição de sorotipos (*serotype-shifting*), já foi relatado em países que adotaram a vacina conjugada (Frazão *et al.*, 2005; Jefferies *et al.*, 2011; Pichon *et al.*, 2013; Rudolph *et al.*, 2013). Assim, é fundamental um monitoramento eficiente dos sorotipos prevalentes no Brasil para que estes possam ser incluídos em futuras versões da vacina contribuindo para a eficácia do programa de imunização anti-pneumococos no país.

Tabela 1: Classificação de *Streptococcus pneumoniae* em sorogrupos e sorotipos.

Sorogrupos	Sorotipos
	Tipo 1
	Tipo 2
	Tipo 3
	Tipo 4
	Tipo 5
Grupo 6	6A, 6B, 6C, 6D
Grupo 7	7F, 7A, 7B, 7C
	Tipo 8
Grupo 9	9A, 9L, 9N, 9V
Grupo 10	10F, 10A, 10B, 10C
Grupo 11	11F, 11A, 11B, 11C, 11D
Grupo 12	12F, 12A, 12B
	Tipo 13
	Tipo 14
Grupo 15	15F, 15A, 15B, 15C
Grupo 16	16F, 16A
Grupo 17	17F, 17A
Grupo 18	18F, 18A, 18B, 18C
Grupo 19	19F, 19A, 19B, 19C
	Tipo 20
	Tipo 21
Grupo 22	22F, 22A
Grupo 23	23F, 23A, 23B
Grupo 24	24F, 24A, 24B
Grupo 25	25F, 25A
	Tipo 27
Grupo 28	28F, 28A
	Tipo 29
	Tipo 31
Grupo 32	32F, 32A
Grupo 33	33F, 33A, 33B, 33C, 33D
	Tipo 34
Grupo 35	35F, 35A, 35B, 35C
	Tipo 36
	Tipo 37
	Tipo 38
	Tipo 39
	Tipo 40
Grupo 41	41F, 41A
	Tipo 42
	Tipo 43
	Tipo 44
	Tipo 45
	Tipo 46
Grupo 47	47F, 47A
	Tipo 48

Informação extraída de: STATENS SERUM INSTITUT. *Pneumococcal antisera*. Disponível em: < <http://www.ssi.dk/> >. Acesso em: 15 Jan. 2014.

1.4 *S. pneumoniae* e o *locus cps*

As cápsulas são estruturas de proteção presentes na superfície da maioria das bactérias. A cápsula polissacarídica atua como uma barreira inespecífica, diminuindo a deposição do complemento na superfície bacteriana e, assim, limitando a interação desta com os fagócitos. A expressão dos polissacarídeos capsulares (CPS) é essencial para a sobrevivência do pneumococo no sangue de pacientes infectados e, portanto, para sua virulência, mas também é um importante alvo para os anticorpos do hospedeiro e para o desenvolvimento de vacinas. Cada sorotipo de pneumococo apresenta uma composição diferente em sua CPS (Henrichsen, 1995). Atualmente, são conhecidos 92 sorotipos de pneumococos caracterizados pelos padrões de reatividade observados contra antissoros específicos que reconhecem variantes antigênicas nos polissacarídeos capsulares sorotipo-específicos expressos na superfície dos pneumococos (Park *et al.*, 2007; Bratcher *et al.*, 2010), mas nem todos os tipos capsulares conhecidos podem ser associados a doenças em humanos. Menos que 30 sorotipos são responsáveis por mais de 90% dos casos de doenças invasivas (Whitney *et al.*, 2003; Martens *et al.*, 2004). Na América Latina, 13 sorotipos são responsáveis por mais de 85% dos isolados invasivos (Dias *et al.*, 2007). Acredita-se que esta diversidade substancial encontrada nos CPSs dos pneumococos tenha surgido como consequência da forte pressão seletiva exercida pelo sistema imune humano ao longo do tempo (Bentley *et al.*, 2006).

Os 92 sorotipos são divididos em 21 grupos e 25 tipos, sendo que cada grupo é constituído por 2-5 sorotipos que apresentam semelhanças imunológicas suficientes para serem classificados no mesmo grupo (Mavroidi *et al.*, 2007) (Tabela 1). A maioria dos genes que codificam as enzimas envolvidas na síntese das CPS estão localizados no mesmo *locus cromossômico*, denominado *cps*, em todos os diferentes sorotipos já identificados. Outros estreptococos como *oralis* e *mitis* também possuem os genes que codificam as enzimas para síntese de seu CPS localizados no *locus cps*. A maioria das CPS pneumocócicas são geralmente sintetizadas pela via *Wzx/Wzy*-dependente, exceto para os sorotipos 3 e 37 que usam a via da sintase. As enzimas responsáveis pela síntese da CPS em pneumococos são codificadas por um conjunto de genes localizados no *loci cps*, que é flanqueado por *dexB* (*glucan 1,6-alpha-glucosidase*) e *aliA* (*oligopeptide ABC transporter substrate-binding protein*) (Bentley *et al.*, 2006). Estes genes codificam transferases, polimerases, flipases, complexo de translocases *Wzd/Wze*, transposases, entre outras.

O sequenciamento do *loci cps* dos 90 sorotipos de pneumococos publicados em 2006 permitiu a realização de análises genéticas, químicas e imunológicas para prever o papel dos

genes *cps* na síntese do CPS. Os resultados das análises revelaram que diferentes mecanismos contribuem para a diversidade antigênica do CPS. Alguns destes mecanismos estão envolvidos na divergência de um único sorotipo em dois outros pelo acúmulo de mutações pontuais (sorogrupo 6), ou a inserção ou deleção de um único gene, resultando em leves mudanças estruturais no CPS (sorogrupo 18). Em outros casos, o *loci cps* de alguns sorotipos dentro do mesmo sorogrupo parecem ter evoluído de forma independente e provavelmente refletem o compartilhamento de um epítipo dominante que os agrupa no interior do mesmo sorogrupo (por exemplo, sorogrupos 7, 17, 33 e 35). Da mesma forma, alguns sorotipos colocados em diferentes sorogrupos mostram mais parentesco entre os seus *loci cps* que aqueles dentro do mesmo sorogrupo (por exemplo, tipos 7B e 7C estão mais estreitamente relacionados com o tipo 40 que com 7A e 7F). Isto se explica pelo fato dos sorogrupos terem sido definidos pela presença de epítipos comuns na ausência de qualquer conhecimento das estruturas dos CPSs ou das sequências dos genes *cps* que codificam sua síntese. As cepas que compartilham um epítipo comum são classificadas dentro do mesmo sorogrupo, mesmo que existam grandes diferenças em outras partes da estrutura do CPS e, portanto, na sequência do *cps*. As sequências *cps*, associadas as estruturas dos polissacarídeos e aos perfis sorológicos, constituem um extenso conjunto de dados que propiciaram avanços na sorotipagem molecular dos pneumococos (Bentley *et al.*, 2006; Aanensen *et al.*, 2007; Mavroidi *et al.*, 2007).

A biossíntese e montagem do polissacarídeo capsular são processos complexos. Modelos hipotéticos da biossíntese capsular baseados numa mistura de evidências experimentais e especulações têm sido propostos (Bentley *et al.*, 2006). Resumidamente, nesse modelo proposto, a cápsula é sintetizada em etapas sequenciais. Na primeira etapa, acontece a transferência de um monossacarídeo fosfato inicial de um nucleotídeo açúcar difosfato presente no citoplasma para um carreador lipídico associado à membrana citoplasmática por enzimas associadas à membrana interna. Numa segunda, ocorre a adição sequencial de monossacarídeos do citoplasma para o polissacarídeo nascente por enzimas associadas à membrana interna para produzir a unidade repetida da cápsula que será ligada ao carreador lipídico associado à membrana citoplasmática. A unidade repetida recém-formada é, então, movida através da membrana citoplasmática para o lado externo da célula por um transportador de unidade repetida, ou flipase. Uma polimerase forma o CPS maduro adicionando novas unidades repetidas e, posteriormente, a CPS formada é ligada ao peptideoglicano por um complexo enzimático (Figura 1).

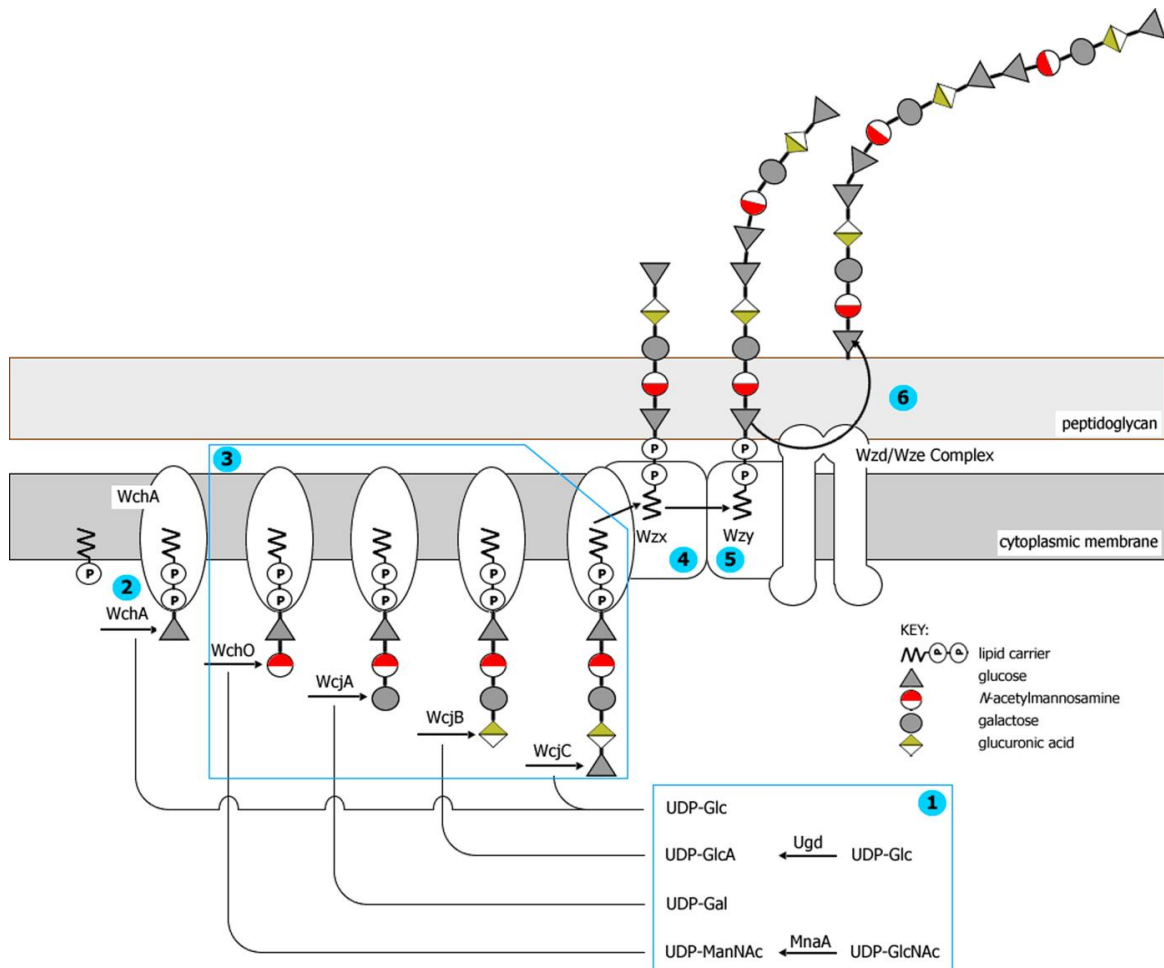


Figura 1: Representação da via *dependente Wzx/Wzy* para biossíntese do CPS 9A.

Esta figura é um modelo hipotético da biossíntese capsular em *S. pneumoniae* baseado na combinação de evidências experimentais e especulações.

- (1) Biossínteses de açúcar nucleotídeo *non-housekeeping*.
- (2) A transferase inicial (neste caso *WchA*) liga o açúcar inicial com um açúcar fosfato (Glc-P) a um carreador lipídico associado a membrana (amplamente considerados com fosfato undecaprenil).
- (3) Glicosil transferases ligam, sequencialmente, mais açúcares para gerar a unidade repetida.
- (4) *Wzx* flipase transporta a unidade repetida através da membrana citoplasmática.
- (5) *Wzy* polimerase liga unidades repetidas individuais para formar a CPS ancorada a lipídeo.
- (6) Complexo de translocases *Wzd/Wze* transporta a CPS madura para a superfície da célula e também é responsável por ancora-la ao peptideoglicano.

Figura extraída de (Bentley *et al.*, 2006).

Uma característica marcante do *locus cps* é a presença de muitas formas altamente divergentes de cada uma das principais classes de enzimas. Assim, há 40 grupos homólogos para polimerases de polissacarídeos, 13 grupos de flipases, e uma grande diversidade de transferases. A presença de múltiplas formas destas enzimas não homólogas, ou altamente divergentes, em conjunto com o baixo percentual de conteúdo G+C da região em que estas enzimas são codificadas, apoia a ideia de que estes genes foram importados pelos pneumococos (ou seus ancestrais) em múltiplas ocasiões de diferentes fontes desconhecidas, o que se reflete na diversidade química da cápsula do pneumococo (Bentley *et al.*, 2006; Mavroidi *et al.*, 2007).

O *locus cps* é flanqueado por dois genes com alto grau de conservação, o *dexB* e *aliA* (García *et al.*, 2000). Bentley e colaboradores (2006) amplificaram o *locus cps* de 90 sorotipos de *S. pneumoniae* com o uso de iniciadores complementares aos genes *dexB* e *aliA*. As ampliações geraram produtos com tamanho variando entre 10.337 e 30.298 pb (tamanho médio = 20.714 pb) que foram, em seguida, sequenciados. As análises das sequências obtidas revelaram que os genes envolvidos na síntese da CPS estão sempre orientados na mesma direção de *dexB* para *aliA*, e que alguns genes do *locus cps* apresentam um alto grau de conservação entre os diferentes sorotipos, como é o caso dos quatro genes *wzg*, *wzh*, *wzd* e *wze* (ou *cps ABCD*), enquanto outros genes, como o *wzx* e *wzy*, apresentam maior diversidade, além de um conteúdo G+C menor que o dos genes *cps ABCD* (Figura 2). Esses mesmos autores destacam que, na região entre os genes que codificam o CPS e os genes que os flanqueiam, há indícios da presença de elementos móveis em quase todos os sorotipos. Ainda não foi totalmente esclarecido como essas alterações genéticas do *locus cps* foram sendo incorporadas ao longo do tempo.

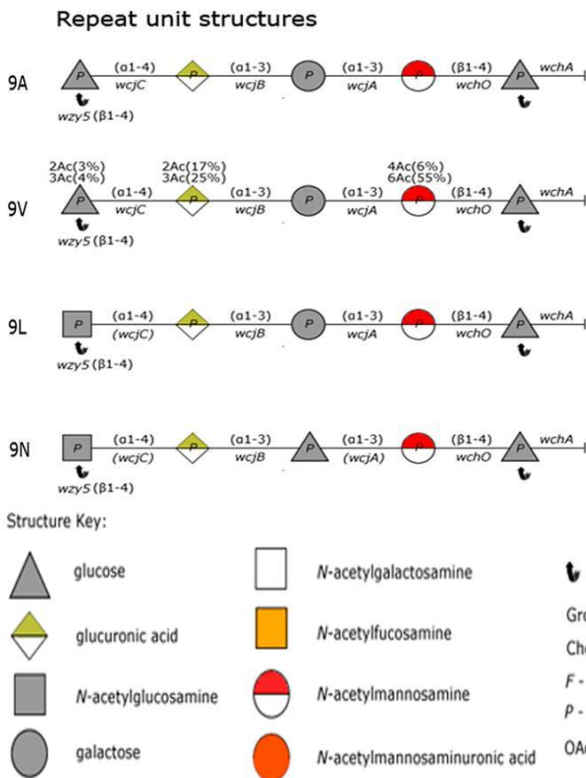
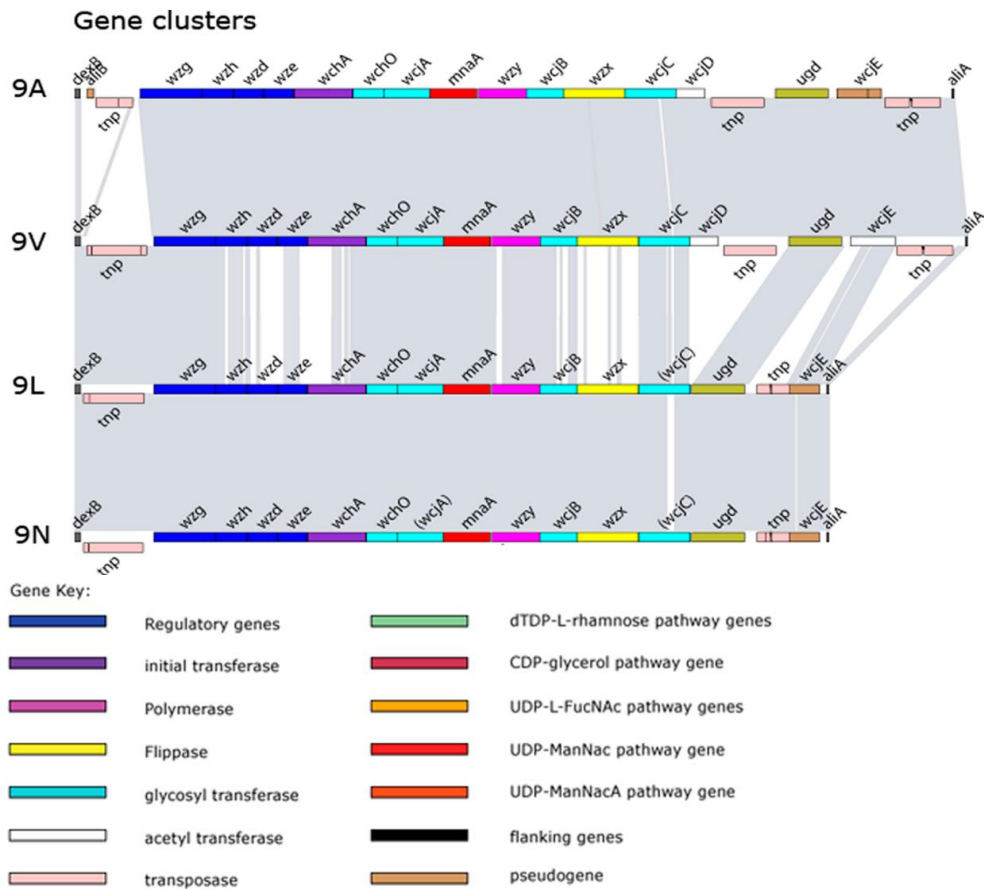


Figura 2: Representação esquemática de alguns clusters CPS e das respectivas unidades repetidas.

Figura extraída de (Bentley *et al.*, 2006).

1.5 Sorotipagem de *S. pneumoniae*

Já em 1902, o cientista Friedrich Neufeld, na Alemanha, noticiou que quando os pneumococos eram misturados a antissoros homólogos imunes ocorria um acentuado inchaço (*Quellung*) na cápsula (Neufeld, 1902). Em 1931 e 1932, Armstrong, na Inglaterra, relatou seus resultados de tipagem de pneumococos em escarro e exsudato peritoneal de ratos usando um método baseado neste fenômeno (Beckler e Macleod, 1934). Beckler e Macleod (1934) ampliaram o método já usado para a determinação de tipos I, II e III, para a determinação dos outros 29 tipos específicos de pneumococos. Essa técnica era utilizada para tipagem de pneumococos em escarros, *pools* de antissoros de A a H eram usados para fazer uma triagem inicial e, posteriormente, os antissoros monovalentes individuais dos tipos I ao XXXII eram utilizados para chegar ao sorotipo específico. Statens Serum Institut (Copenhague, Dinamarca) divulgou, em 1963, a produção de *kits* para sorotipagem de pneumococos.

Os *kits* eram compostos de nove *pools* de soros, com sete a 11 tipos cada, e 81 soros univalentes para identificação dos 81 sorotipos conhecidos (Lund, 1963). Erna Lund, em 1970, propôs alterações na nomenclatura da classificação de pneumococos em tipos e recomendou a nova classificação como sistema internacional de tipagem (Lund, 1970). Em 1978, Russell *et al.* (1978) relataram que o método clássico chamado de reação de *Neufeld*, ou, de *Quellung*, já era capaz de identificar 83 sorotipos divididos em 15 grupos e 15 tipos.

Em 1985, Austrian *et al.* (1985) descreveram um novo tipo: 16A. Um total de 84 tipos de pneumococos foi descrito até 1993, os tipos que exibiam estreita reatividade sorológica cruzada eram agrupados. Dos 84 tipos, 58 pertenciam a 20 grupos contendo de dois a quatro tipos, e um total de 46 diferentes tipos ou grupos era conhecido. Soros fatores monovalentes tornados mais específicos por adsorções múltiplas ou por indução de tolerância imunológica a uma reação cruzada eram utilizados para identificar os tipos dentro dos grupos (Sørensen, 1993). Sorensen (1993) descreveu um novo sistema de tipagem de pneumococos (*chessboard*), baseado em 12 *pools* de antissoros em substituição ao sistema antigo que utilizava nove *pools*, cada *pool* reagindo com 7 a 11 tipos individuais. O *chessboard* era destinado à tipagem de um maior número de pneumococos isolados de sangue e líquido cefalorraquidiano (Figura 3). Em 1995, a propriedade sorológica de seis novos sorotipos foi descrita (Henrichsen, 1995), totalizando 90 sorotipos. Park *et al.* (2007) identificaram um novo sorotipo entre *Streptococcus pneumoniae* do grupo 6, que deram o nome de sorotipo 6C, e, em 2010, foi descrito o sorotipo 6D, surgido após o gene *wciN* do locus *cps* do sorotipo 6A

ter sido substituído por um novo gene, totalizando os 92 sorotipos hoje conhecidos (Bratcher *et al.*, 2010).

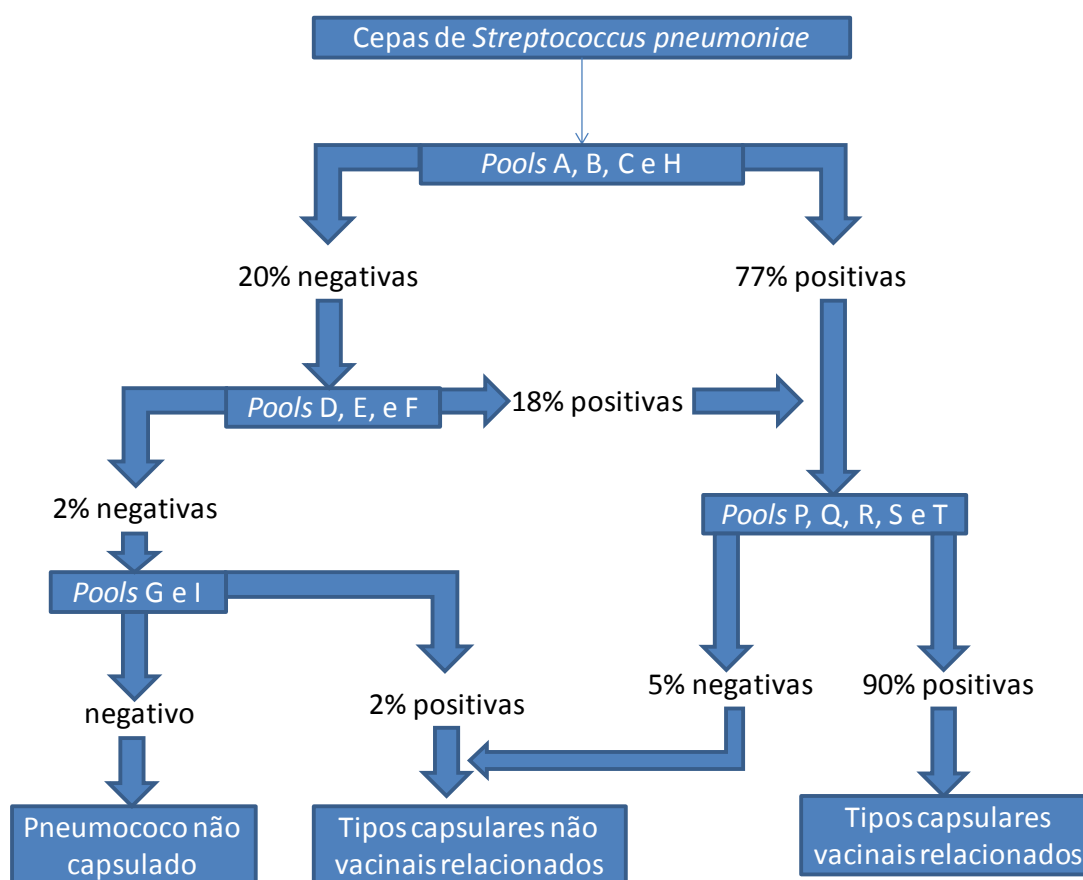


Figura 3: Fluxograma mostrando a estratégia para sorotipagem clássica de pneumococos (*chessboard*).

Mais que 90% das cepas isoladas de sangue e líquido podem ser tipadas ou grupadas pelo uso dos 12 *pools* de antissoros (*pools A a F plus H, e pools P a T*) em um sistema (*chessboard*).

Figura extraída de (Sørensen, 1993).

A sorotipagem clássica é o método padrão-ouro para tipagem de *S. pneumoniae* e de outras bactérias cápsuladas como *Klebsiella pneumoniae*, *Haemophilus influenzae*, e *Neisseria meningitidis*. O método atual de sorotipagem de pneumococos utiliza antissoros produzidos por imunização de coelhos com células bacterianas totais mortas, adquiridos do Statens Serum Institut. O painel de antissoros é composto por quatorze *pools* que contêm antissoros contra oito a 14 tipos, designados de A a T, e 21 antissoros grupos específicos, 25 tipos específicos, 64 fatores de antissoros (Henrichsen e Robbins, 1992) para identificar os sorotipos dentro dos grupos e um antissoro *Anti C-polysaccharide* para identificar

pneumococos não-capsulados. Assim, um total de 126 antissoros é requerido para tipagem de todos os 92 sorotipos de acordo o Statens Serum Institut.

Nessas reações capsulares, quantidades iguais de culturas bacterianas e antissoros são misturadas em lâminas de vidro, e a reação é examinada em microscópio de contraste de fase, sendo que uma reação positiva é identificada quando a cápsula que envolve o pneumococo incha e torna-se visível. As bactérias, frequentemente, também aglutinam. Esse método é caro, trabalhoso, apresenta reações cruzadas devido às semelhanças entre os antígenos presentes na cápsula de sorotipos do mesmo sorogrupo, ou mesmo de sorogrupos diferentes, e necessita do uso de animais para a produção de antissoros, indo na contra mão das pressões sociais crescentes para redução dos experimentos com animais (Lalitha *et al.*, 1996; Mavroidi *et al.*, 2007). Além disso, o método não é capaz de detectar sorotipos ainda não descritos, para os quais não existem antissoros. Outro problema decorrente da utilização do método clássico de sorotipagem de pneumococos é a demora na obtenção dos resultados, pois poucos são os laboratórios certificados pela OMS para prestação deste serviço (Vieira *et al.*, 2007).

No Brasil, a sorotipagem de *S. pneumoniae* é centralizada no Instituto Adolf Lutz – São Paulo, centro de referência para meningites bacterianas no país. Devido à grande demanda, os resultados podem demorar até um ano para serem remetidos aos solicitantes. Além disso, alguns isolados deixam de produzir a cápsula após repiques sucessivos ou longo tempo de armazenamento, o que os torna não-tipáveis à sorotipagem clássica.

Métodos alternativos à reação de *Quellung* têm sido descritos, como, por exemplo, a reação de contraímunoeletroforese (CIE), o teste de aglutinação do látex (LA), *Pneumotest-Latex*, teste de coaglutinação (COA), *Dot blot assay*, *fluorescent antibody technique*, *capillary precipitin test*, *Multibead Assay*. Todos estes métodos são baseados em reações antígeno-anticorpo (Kronvall, 1973; Holliday, 1981; Lalitha *et al.*, 1996; Lalitha *et al.*, 1999; Park, Briles e Nahm, 2000; Mudany *et al.*, 2003; Shutt, Samore e Carroll, 2004; Slotved *et al.*, 2004; Sanz *et al.*, 2010).

1.6 Sorotipagem molecular

Métodos alternativos de tipagem não baseados na detecção do polissacarídeo capsular têm sido descritos para *S. pneumoniae*. Esses métodos utilizam técnicas de biologia molecular ou fenotípicas, incluindo o *pulsed-field gel electrophoresis (PFGE)*, *restriction fragment end labeling*, *BOX PCR*, *arbitrarily primed PCR*, *multilocus sequence typing (MLST)*, *Multiplex PCR*, *Real Time PCR*, *PCR e Electrospray Ionization Mass Spectrometry* (Shaaly *et al.*, 2005; Pai, Gertz e Beall, 2006; Hathaway *et al.*, 2007; Iraurgui *et al.*, 2010; Miernyk *et al.*, 2011;

Yu *et al.*, 2011; Yun *et al.*, 2011; Massire *et al.*, 2012). Embora estes métodos sejam úteis para distinguir clones em estudos epidemiológicos, seus resultados não se correlacionam com os da sorotipagem que permanece sendo o padrão ouro para tipagem bacteriana, devido tanto à importância da cápsula na patogênese e imunidade, quanto à riqueza de dados epidemiológicos acumulados com o uso dessa técnica secular.

Após a caracterização da região *cps* de 90 sorotipos ocorreu a publicação de novos ensaios para sorotipagem molecular de pneumococos, como por exemplo, multiplex PCR sequencial (smPCR), multiplex PCR (mPCR) combinado com análises de fragmentos por eletroforese capilar fluorescente automatizada (FAF-mPCR), *PCR-based automated microarray assay*, *Triplex Real-Time PCR*, entre outros. Estes ensaios baseiam-se na detecção de genes, ou conjunto de genes que caracterizam um sorotipo específico. Os ensaios de multiplex PCR estão cada vez mais sendo utilizados devido ao seu potencial de ultrapassar algumas dificuldades relatadas nos testes sorológicos e por oferecerem a oportunidade de serem adaptados a áreas geográficas específicas, mas apesar desta técnica ser de fácil execução, o conjunto de sorotipos que podem ser identificados é limitado (Pai, Gertz e Beall, 2006; Dias *et al.*, 2007; Morais *et al.*, 2007; Jourdain *et al.*, 2011; Vickers *et al.*, 2011; Yun *et al.*, 2011; Coskun-Ari, Guldemir e Durmaz, 2012; Selva *et al.*, 2012; Collard, Alio Sanda e Jusot, 2013; Raymond *et al.*, 2013).

Em 1999, foi desenvolvido um método alternativo para sorotipagem molecular de *Escherichia coli* – *Shigella*. O método, designado *rfb*-RFLP (Coimbra, Grimont e Grimont, 1999; Coimbra *et al.*, 2000), mostrou-se mais robusto que a sorotipagem tradicionalmente feita para essas bactérias, também realizada por meio de antissoros (sorotipagem clássica). Através do *rfb*-RFLP, Coimbra e colaboradores (Coimbra *et al.*, 2001), e outros pesquisadores (Melito, P. L. *et al.*, 2005; Grimont, F. *et al.*, 2007), descreveram novos sorotipos de *Shigella* e *E. coli* previamente não-tipáveis à sorotipagem clássica. A técnica consiste na amplificação de uma região do DNA conhecida como *rfb* que codifica os genes, em “cluster”, para síntese do antígeno-O (antígeno capsular somático). A amplificação dessa região nos vários sorotipos empregando-se um único par de iniciadores é possível devido à conservação dos genes que flanqueiam a região *rfb*. O produto da amplificação é subsequentemente digerido com a enzima de restrição MboII gerando perfis únicos (*fingerprints*) à eletroforese, semelhante a um código de barras, e permitindo assim a diferenciação de cada sorotipo.

Recentemente, Coimbra e colaboradores (Coimbra *et al.*, 2010), desenvolveram um *software web-based* de acesso livre que permite identificar automaticamente os sorotipos de *Shigella* - *E. coli* a partir de seus *fingerprints* obtidos com o método do *rfb*-RFLP. O *software*

denominado MST (“*Molecular Serotyping Tool*”) pode ser acessado nos servidores do Centro de Excelência em Bioinformática (CEBio) do Centro de Pesquisa René Rachou, FIOCRUZ (<http://www.cebio.org/mst>). Para usá-lo, basta que o usuário insira, no campo indicado, o tamanho em pares de bases dos fragmentos de restrição correspondentes as bandas observadas na eletroforese usando a técnica padronizada (Figura 4). O algoritmo de programação dinâmica do MST é inspirado no algoritmo de Needleman-Wunsch para alinhamento global de sequências (Needleman e Wunsch, 1970).

MST considera três operações de edição capazes de transformar um perfil em outro: a) inserção de uma banda em um dos perfis; b) deleção de uma banda em um dos perfis; e c) substituição de uma banda por outra de tamanho diferente até um limite aceitável. Essa tolerância é determinada a partir da observação prévia das variações intra- e inter-gel dos pesos moleculares dos fragmentos calculados a partir de suas distâncias de migração a eletroforese. Uma equação linear define o percentual de erro aceitável em cada faixa de tamanho dos fragmentos, uma vez que as bandas menores migram mais e tendem a ser mais difusas que os fragmentos maiores. A maior difusão dos fragmentos pequenos compromete a acurácia na estimativa de seu tamanho. Este erro não pode exceder a penalidade para uma deleção de banda, que, é definido por um limiar variável, correspondendo ao erro máximo do dimensionamento linear do fragmento de 7,0% em 0,5 kpb a 3,5% em 4 kpb. A função escore assinala uma penalidade para cada transformação, e o algoritmo de programação dinâmica computa o melhor escore de edição entre dois padrões produzindo uma matriz de edição. O escore global (Ge) é calculado usando o escore de edição (operações de edição recebem penalidades e sua somatória produz o escore de edição), o número de fragmentos não alinhados e o total do número de fragmentos nos dois padrões ($Ge = \text{escore de edição} \times \text{número de fragmentos não alinhados} \times 100 / (\text{número de bandas})^2$). Ge é mais influenciado por falta de correspondência entre duas bandas idênticas que por erros no dimensionamento das bandas, que é o mais comum artefato que afeta a reprodutibilidade dos métodos à base de RFLP. O programa compara o perfil fornecido pelo usuário contra todos os perfis de referência do banco, retornando ao usuário uma representação esquemática do alinhamento do perfil desconhecido contra o perfil do sorotipo mais semelhante do banco (menor escore) (Mond, Lees e Snapper, 1995; Coimbra *et al.*, 2010) (Figura 5).

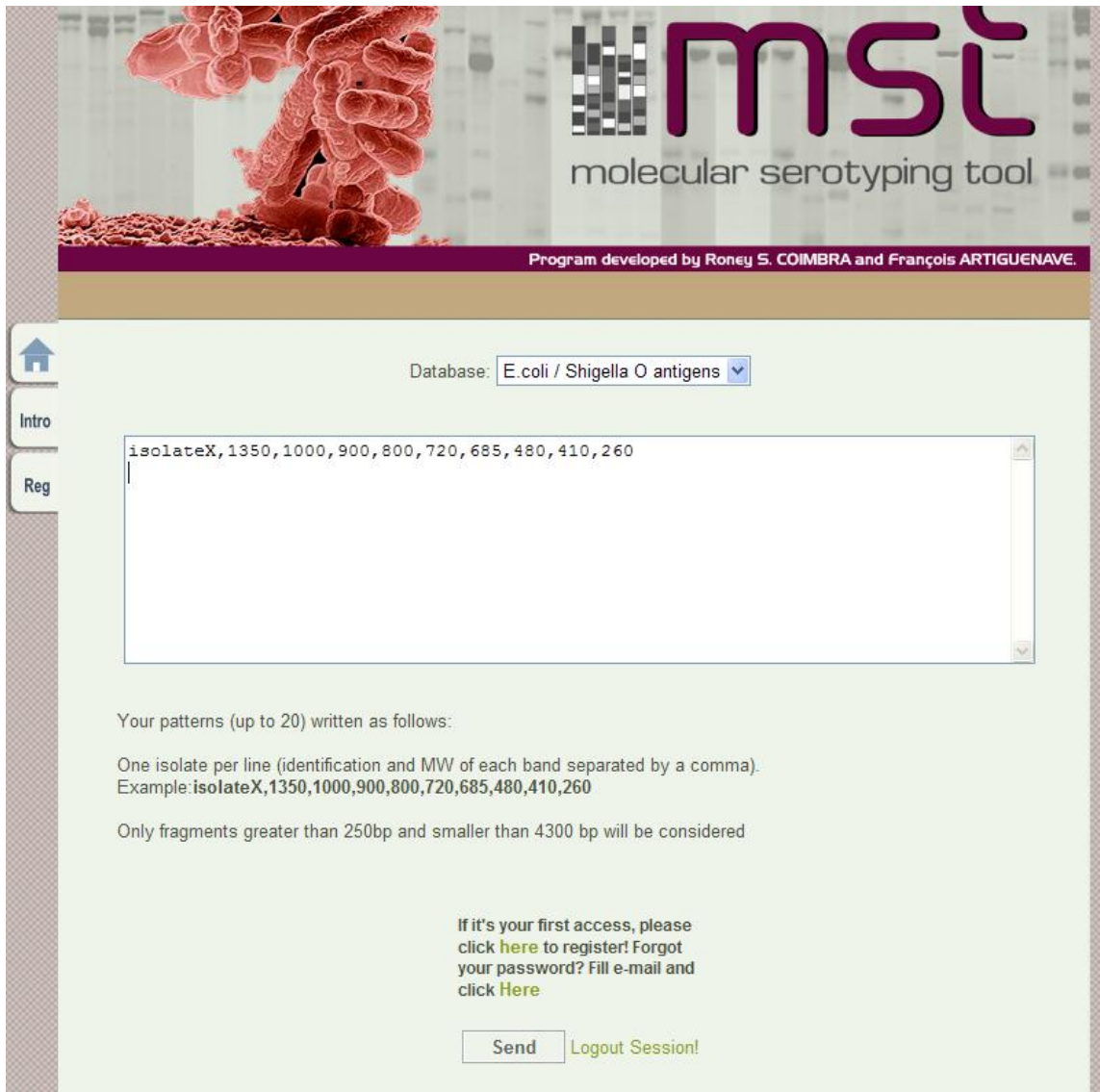
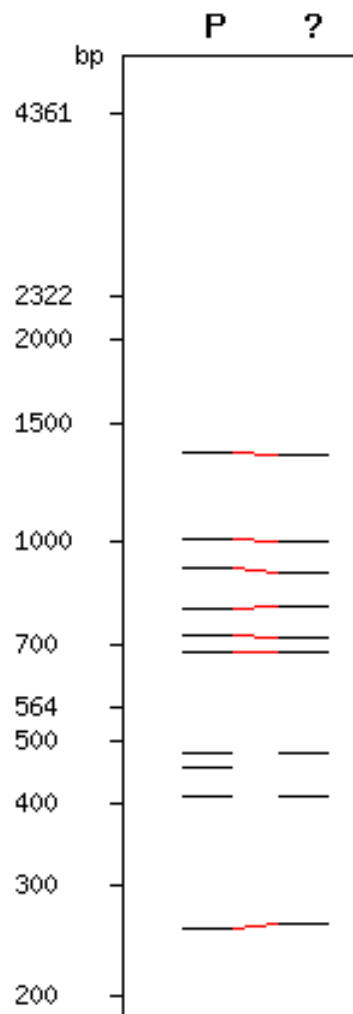


Figura 4: *Molecular Serotyping Tool* for *E. coli/Shigella spp* acessível nos servidores do CEBio (<http://www.cebio.org/mst>).

O perfil *rfb*-RFLP obtido na etapa experimental é introduzido no software MST e a identificação do sorotipo é solicitada clicando no botão “*Send*”.

Schematic view



Experimental pattern (?): isolateX
Closest reference pattern (P): A1
Database: E.coli_Shigella
Score = 0.3

Figura 5: Resultado esquemático após análise do perfil experimental pelo Software MST.

O software MST (*"Molecular Serotyping Tool"*) retorna ao usuário uma representação esquemática do alinhamento do perfil desconhecido contra o perfil de referência mais semelhante contido no banco e o menor valor de escore dos alinhamentos.

Legenda: **P:** Perfil do padrão contido no banco de dados do software MST; **?:** Perfil desconhecido do isolado clínico; **BP:** Pares de bases; **Linha vermelha:** indica o alinhamento entre bandas correspondentes nos dois perfis.

O desenvolvimento desta tecnologia gerou um enorme avanço na sorotipagem das enterobactérias tendo possibilitado, inclusive, a identificação de novos sorotipos (Coimbra, R. *et al.*, 2001; Brisse, Issenhuth-Jeanjean e Grimont, 2004; Melito, P. *et al.*, 2005; Grimont, F *et al.*, 2007). Posteriormente, o desenvolvimento do software MST e sua disponibilização livre à comunidade científica através de um servidor *web-based* possibilitou que pesquisadores do mundo inteiro pudessem comparar os perfis *rfb*-RFLP obtidos em seus laboratórios com os perfis de referência depositados no banco que integra o programa (Coimbra *et al.*, 2010).

Fica evidente, portanto, a necessidade do desenvolvimento de novos métodos alternativos a reação de *Quellung* que possibilitem que mais Laboratórios de Saúde Pública realizem a sorotipagem de pneumococos, contribuindo, dessa forma, para o aperfeiçoamento da vigilância epidemiológica dos sorotipos circulantes no país após a introdução da vacina pneumocócica decavalente no calendário básico de vacinação.

1.7 Justificativa

O grande número de casos de doenças pneumocócicas no Brasil levou o Ministério da Saúde a incluir a vacina pneumocócica conjugada decavalente no programa nacional de imunização em 2010. Alguns estudos realizados após a introdução da vacina conjugada em diferentes comunidades de países desenvolvidos mostraram que houve um aumento na incidência de sorotipos que não estavam incluídos na vacina, fenômeno chamado de *serotype-shifting* (Pletz *et al.*, 2008). Conhecer a distribuição dos sorotipos é, portanto, pré-requisito para se monitorar a eficácia dos programas de vacinação contra pneumococos. Atualmente, a sorotipagem dos isolados clínicos de pneumococos só é possível em centros que detêm a coleção completa de antissoros contra todos os sorotipos conhecidos. Diante disso, fez-se necessário o desenvolvimento de uma metodologia para sorotipagem molecular que seja mais acessível, eficiente, rápida e de baixo custo. Outras vantagens da sorotipagem molecular incluem a dispensa do uso de animais para a produção dos antissoros e, também, a possibilidade de detecção de sorotipos ainda não descritos para os quais, conseqüentemente, não há antissoros disponíveis. A participação da Funed no desenvolvimento do método *cps*-RFLP e a incorporação, em sua rotina, deste novo método acoplado ao software e bando de dados MST servirá como “prova de conceito” para que, no futuro, o sistema possa ser expandido para os outros Laboratórios de Referência da federação permitindo um monitoramento mais eficiente dos sorotipos de pneumococos prevalentes no Brasil e orientando as adaptações que a vacina conjugada pneumocócica deverá sofrer para manter-se eficiente ao longo do tempo.

O Estudo de Viabilidade Técnica, Econômica, Comercial e de Impactos Ambientais e Sociais (EVTECIAS) realizado pelo Núcleo de Tecnologia da Qualidade da Inovação (NTQI) no âmbito do Programa de Incentivo à Inovação (PII) da FIOCRUZ-Minas concluiu que o método *cps-RFLP*, para sorotipagem molecular, acoplado ao software e banco de dados MST, para identificação dos sorotipos, é um método viável e de grande interesse para o sistema público de saúde brasileiro (FIOCRUZ, 2011).

2.1 Objetivo Geral

Desenvolver e padronizar uma nova ferramenta de sorotipagem molecular para *Streptococcus pneumoniae* baseada no polimorfismo dos tamanhos dos fragmentos de restrição do produto da amplificação do *loci cps* (*cps*-RFLP) que codifica as enzimas responsáveis pela biossíntese da CPS nesta espécie.

2.2 Objetivos específicos

- 1- Selecionar por meio de análises *in silico* a enzima de restrição mais discriminante para utilização no método de sorotipagem molecular para pneumococos;
- 2- Construir um banco de dados com os perfis de referência obtidos *in silico* após digestão com a endonuclease mais discriminante (*cps*-RFLP) e posteriormente carregar o software MST com o banco de dados *in silico*;
- 3- Padronizar a metodologia experimental e obter os perfis *cps*-RFLP de um painel de amostras clínicas representativas dos sorotipos prevalentes em Minas Gerais;
- 4- Complementar o banco de perfis *cps*-RFLP do software MST com os perfis experimentais, que eventualmente difiram dos perfis calculados *in silico*.

3 MATERIAL E MÉTODOS

3.1 Local do estudo

A pesquisa foi conduzida no Laboratório de Parasitologia Celular e Molecular (LPCM) – FIOCRUZ-Minas, Centro de Excelência em Bioinformática (CEBio) – FIOCRUZ-Minas e no Serviço de Doenças Bacterianas e Fúngicas da Divisão de Epidemiologia e Controle de Doenças do Instituto Octávio Magalhães da Fundação Ezequiel Dias - Laboratório Central de Saúde Pública de Minas Gerais (LACEN-MG).

3.2 Sequências de DNA da região *cps* de *S. pneumoniae*

As regiões *cps* de 91 cepas, representando 90 sorotipos de *Streptococcus pneumoniae*, anteriormente publicadas foram analisadas utilizando-se sequências de DNA depositadas no *GenBank* (<http://www.ncbi.nlm.nih.gov/genbank>), cujos identificadores foram CR931632 a CR931722 (Bentley *et al.*, 2006).

3.3 Procedimentos experimentais

3.3.1 Análises *in silico*

O arquivo multi-FASTA que continha as sequências de DNA da região *cps* dos 90 sorotipos de pneumococos foi utilizado para pesquisar quais as enzimas de restrição possuíam sítios de clivagem nas sequências analisadas usando-se o software REMAP, do pacote *European Molecular Biology Open Software Suite* (EMBOSS). Um *script* em linguagem de programação Perl (*Practical Extraction and Report Language*) foi utilizado para identificar, no arquivo gerado pelo REMAP, quais as enzimas de restrição que apresentavam entre quatro e 26 sítios de clivagem em todas as 90 sequências de DNA analisadas. Para cada enzima selecionada, criamos um banco de dados com os padrões de restrição *in silico* previstos referentes à região *cps* de cada um dos 90 sorotipos gerados com o software RESTRICT, do pacote EMBOSS. Utilizamos outro *script* em linguagem de programação Perl para capturar o identificador de cada sequência e o tamanho dos fragmentos gerados ordenados do maior para o menor, sendo que o *script* também descartava as bandas inferiores a 250 pb e superiores a 4300 pb. Os arquivos *.mst gerados foram utilizados para produzir as matrizes de distância par-a-par entre todos os perfis. Para isso, empregou-se o software *Molecular Serotyping Tool* (MST) (Coimbra *et al.*, 2010). O MST calcula os escores de distância entre os perfis, definidos como a soma das penalidades para a operação de edição que transforma um padrão

no outro. Para a etapa de validação experimental foi selecionada a enzima de restrição que produziu o conjunto de *fingerprints* mais diversos. Essa enzima, sendo a mais discriminante (maiores escores nos alinhamentos par-a-par, ou seja, aquela que produziu a maior média de distância calculada e o menor número de reações cruzadas) permitiu a diferenciação entre o maior número de sorotipos. Foi utilizado o teste de *Wilcoxon rank-sum* para comparação dos resultados obtidos com as diferentes enzimas de restrição (vide tópico abaixo).

O software NEIGHBOR, do pacote PHYLIP (PHYLogeny Inference Package), com o UPGMA (*Unweighted Pair Group Method with Arithmetic Mean*) como método de ligação, foi usado para construir as árvores a partir das matrizes de distância geradas pelo software MST. As árvores foram visualizadas e editadas com o software livre FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Os scripts em linguagem de programação Perl desenvolvidos pelo Centro de Excelência em Bioinformática do Centro de Pesquisas René Rachou utilizados no âmbito deste estudo estão descritos nos ANEXOS X a XV.

3.3.1.1 Comparação da enzima de restrição selecionada com a endonuclease HinfI

A enzima HinfI já tinha sido proposta para utilização na sorotipagem molecular de um conjunto de 46 sorotipos de pneumococos (Batt *et al.*, 2005). A fim de estimar-se o ganho possível com o método *cps*-RFLP objeto deste estudo, foram comparados o poder discriminatório do *cps*-RFLP empregando-se a enzima selecionada na etapa de análise *in silico* e aquele obtido utilizando-se a enzima HinfI tal como previamente publicado. Para cada uma das enzimas de restrição foi gerado um banco de dados com os padrões *cps*-RFLP previstos por digestão *in silico* das 90 sequências de DNA do *cluster cps*. Os padrões *cps*-RFLP previstos para cada sorotipo foram comparados com o algoritmo de alinhamento de perfis de restrição do *software* MST e os escores dos alinhamentos par-a-par foram registrados. Para identificar-se a enzima mais discriminante, foi utilizada a média, mediana e o desvio padrão dos escores de distâncias calculados entre todos os pares de padrões *cps*-RFLP e o número de pares indistinguíveis para cada enzima. Foi utilizado o teste de *Wilcoxon rank-sum* para comparação dos resultados obtidos com as diferentes enzimas de restrição (vide tópico abaixo).

3.3.2 Padronização do método experimental

Os procedimentos *in vitro* foram padronizados em ensaios com isolados de *S. pneumoniae* previamente sorotipados pela reação de *Quellung* (Sørensen, 1993). Quarenta e nove (49) cepas de *Streptococcus pneumoniae* representando os 34 sorotipos mais prevalentes em Minas Gerais foram disponibilizadas pelo Laboratório de Referência Regional para Meningites Bacterianas da Funed. A relação com a identificação das cepas esta descrita no anexo I. Estes isolados foram preservados pelo método de criopreservação a -70°C em caldo BHI (*Brain Heart Infusion*) contendo 15% de glicerol (agente crioprotetor) adaptado de (Smirnova e Avtushenko, 1988).

3.3.2.1 Identificação e sorotipagem dos *Streptococcus pneumoniae*

As cepas de *S. pneumoniae* foram identificadas após cultivo em ágar sangue de carneiro a 5% (Funed, 2013) por 18 a 24 horas em ambiente de microaerofilia. O isolamento de colônias pequenas, umbilicadas, brilhantes, translúcidas ou mucóides alfa-hemolíticas, revelando ser diplococos Gram positivos lanceolados (DGP) à coloração pelo método de Gram, estabeleceu a suspeita de *Streptococcus pneumoniae*. A distinção com outras bactérias Gram positivas foi realizada com as provas da catalase, hemólise em ágar sangue de carneiro a 5%, prova da optoquina, bile solubilidade e teste de aglutinação do látex para pesquisa de antígenos de pneumococos (Arbique *et al.*, 2004b). A seguir, as cepas foram enviadas para a Seção de Bacteriologia do Instituto Adolfo Lutz (IAL), São Paulo (SP) para controle de qualidade da identificação e sorotipagem clássica. A sorotipagem foi realizada pela reação de *Quellung*, que consiste em uma reação antígeno-anticorpo visualizada à microscopia, com soros obtidos do Statens Serum Institut (Mantese *et al.*, 2003).

3.3.2.2 Cultivo dos isolados e extração de DNA cromossômico

Os reagentes foram fabricados pela Sigma-Aldrich (Saint Louis, MO), exceto quando explicitado no texto.

Os isolados foram cultivados em placas de ágar Mueller-Hinton suplementada com sangue de carneiro a 5% (Funed, 2013) por 18 a 24 horas em ambiente de microaerofilia (5 a 10% de CO²). Assim, amostras de DNA foram preparadas seguindo a metodologia descrita por (Coimbra, Grimont e Grimont, 1999), adaptada para *S. pneumoniae*. A concentração celular foi estimada pela medição da densidade óptica da cultura utilizando turbidímetro (Biomérieux, Marcy l'Etoile, França). Dois mililitros de uma suspensão

bacteriana na escala 3 de Mc Farland, correspondendo a aproximadamente $1,8 \times 10^9$ bactérias foram mantidos em tampão de lavagem (NaCl a 1 M, Tris-HCl a 10 mM, pH 8,0). A suspensão bacteriana foi centrifugada a 4000 rpm por 15 minutos a 4°C. O pellet foi ressuspenso em 10 mL de tampão de lavagem e *vortex* por 15 segundos, seguido por centrifugação a 4000 rpm por 15 minutos a 4°C. O pellet foi ressuspenso em 50 µL de tampão de lavagem, 15 µL de lisozima a 20mg/mL e 3 µL de mutanolisina a 5U/mL. A esta suspensão foram acrescentados 68µl de agarose *low melting* a 2% preparada em tampão TE 0,5X (Tris a 10 mM, pH 8,0; EDTA a 0,1 mM, pH 8,0) previamente fundida e estabilizada junto com a cultura em banho Maria a 41°C por 15min. A mistura foi homogeneizada e deixada no banho a 41°C por 10min. Depositou-se 20 µL da mistura em lâminas envoltas por parafilme para confecção dos plugues. As lâminas com os plugues foram levadas para o freezer por cinco minutos. Após este intervalo, as lâminas foram transferidas do freezer para a temperatura ambiente. Após cinco minutos, os plugues foram transferidos para tubos tipo Falcon de 15 ml (Bacton Dickinson Labware, Franklin Lakes, NJ) contendo 1 mL de solução de lise EC (Tris a 6,0 mM, pH 8,0; NaCl a 1 M; EDTA a 0,1 M, pH 8,0; desoxicolato de sódio a 0,2%; N-lauril sarcosil a 0,5%; 5 µL de mutanolisina (5U/mL); 50 µL de lisozima (20 mg/mL); 16,6 µL de BRIJ 58 a 0,25%; 10 µL de RNase I (50mg/mL). Os tubos foram incubados a 37°C em banho-maria *overnight*. No dia seguinte, a solução de lise EC foi descartada e 1mL de tampão ES (EDTA a 0,5 M, pH 9,0; N-lauril sarcosil a 1%, proteinase K a 0,1 mg/mL foi adicionado a cada tubo. Os tubos foram incubados a 51°C *overnight*. No dia seguinte, o tampão ES foi removido, os plugues foram transferidos para novos tubos *Falcon* de 15 mL e 2 lavagens rápidas com 10ml de tampão TE foram realizadas para a retirada do excesso de detergentes. Os plugues foram lavados mais quatro vezes com 10 mL de Tampão TE 1X sob agitação por 60 minutos. Ao final, os plugues foram armazenados em tubos *Eppendorf* de 1,5 ml com 1 mL de TE 1X a 4°C. A qualidade das extrações foi verificada por eletroforese horizontal em gel de agarose a 0,6% em tampão TBE (Tris a 89 mM, ácido bórico a 89 mM, EDTA a 2,5 mM, pH 8,2) sob 4V por centímetro de distância entre os eletrodos por 1h30min em géis de 14 cm de comprimento. Os tamanhos dos fragmentos de DNA foram estimados usando o marcador de pares de bases *lambda Hind III DNA Ladder* (Promega, Madison, WI). A qualidade do DNA obtido após as extrações foi crítica para o sucesso da etapa de amplificação, tendo em vista que, somente as extrações de DNA que forneceram amostras de alta qualidade resultaram em amplificação da região *cps*.

Os plugues (aproximadamente 20 µL) foram fundidos a 68°C por 15 minutos em 20 µL de água ultrapura livre de DNase imediatamente antes do uso e três microlitros foram usados como *template* de DNA para as reações de amplificação por PCR.

Num segundo momento, o kit de extração *Wizard® Genomic DNA Purification Kit* (Promega, Madison, WI) indicado para extração de fragmentos longos de DNA genômico de bactérias também foi testado (Brisse, Issenhuth-Jeanjean e Grimont, 2004; Batt *et al.*, 2005). O rendimento das extrações, partindo-se de aproximadamente $1,8 \times 10^9$ bactérias, utilizando o kit foi superior ao da extração em plugues de agarose, mas, em várias amostras, o DNA mostrou-se parcialmente degradado impedindo a amplificação da região *cps* em várias amostras testadas.

3.3.2.3 Amplificação do cluster *cps*

Os oligonucleotídeos previamente publicados, AliA₂ (5'-ATG CAG CTA AAG TAG TCG CC-3) e DexB₂ (5'-GAC CGT CGC TTC CTA GTT GT-3') (Batt *et al.*, 2005), respectivamente complementares a *aliA* e *dexB* foram utilizados para amplificar o cluster *cps* em uma reação de PCR longa usando AccuTaq LA DNA polymerase. A reação ótima de PCR continha 3 µL de DNA total extraído e diluído como descrito acima, 32 µL de água ultra-pura, 0,5 µL de AccuTaq LA DNA polymerase a 5U/µL, 5 µL de tampão AccuTaq LA, 2,5 µL de dNTPs (*deoxyribonucleotide triphosphates*) a 10 mM, 1 µL de DMSO (*Dimethyl Sulfoxide*) a 99,9% e 3 µL de cada iniciador a 10 mM, para um volume final de 50 µL (todos de. Adição de betaína ao sistema de amplificação não aumentou a qualidade da reação. O ciclo de reação ideal consistiu de um ciclo inicial de 93°C durante 2 min, 10 ciclos de 93°C durante 15 segundos, 50°C durante 30 segundos, e 68°C durante 20 min, 25 ciclos de 93°C durante 15 segundos, 50°C durante 30 segundos, e 68°C durante 20 min (acréscimo de 15 segundos por ciclo), com um passo de extensão final de 68°C por 10 min. Ao final da reação, as amostras foram mantidas a 4°C no termociclador. Todas as reações foram realizadas num Veriti® 96-Well Thermal Cyclor (Life Technologies, São Paulo, Brasil) usando tubos de 0,2 ml de paredes finas (Axygen, Tewksbury, MA). Os produtos foram mantidos a 4°C até a realização da digestão e permaneceram estáveis por algumas semanas. A qualidade das reações foi verificada por eletroforese horizontal em gel de agarose a 0,6% em tampão TBE (Tris a 89 mM, ácido bórico a 89 mM, EDTA a 2,5 mM, pH 8,2) sob 4V por centímetro de distância entre os eletrodos por 90 min em géis de 14 cm de comprimento. Os tamanhos dos produtos de PCR foram estimados usando o marcador de pares de bases *lambda Hind III DNA Ladder* (Promega, Madison, WI).

3.3.2.4 Condições de restrição

Vinte e cinco microlitros do produto da PCR foram incubados com 10 U da enzima *XhoII*, 5 µL do tampão de digestão e 0,2 µL de albumina sérica bovina (Promega, Madison, WI) por quatro horas a 37°C. Em seguida a enzima foi inativada por aquecimento a 65°C por quinze minutos de acordo instruções do fabricante.

3.3.2.5 Separação dos fragmentos da digestão enzimática por eletroforese convencional

Os produtos de digestão foram submetidos imediatamente à eletroforese em gel de agarose (Invitrogen, São Paulo, Brasil) a 1,5 % com a voltagem de 4,5 V por centímetro de distância entre os eletrodos por 4 horas em TAE 0.5X. A cuba de eletroforese horizontal OWL (Thermo Scientific, Wilmington, DE) de 20 x 25 cm possuía 35,5 cm de distância entre os eletrodos, sendo ideal para aplicações genômicas de alta resolução e corridas longas. O padrão de peso molecular usado nas corridas foi o *GeneRuler 1 kb Plus DNA Ladder* (Thermo Scientific, Wilmington, DE). Após a eletroforese, os géis foram corados por 15 minutos em 500 mL de água Milli-Q contendo 40 µL de brometo de etídio (10 mg/mL) e descorado por duas vezes durante 15 minutos em 500 mL de água destilada. As imagens dos géis foram eletronicamente capturadas utilizando o fotodocumentador MultiDoc-It (UVP, Upland, CA). As imagens capturadas foram armazenadas no formato JPEG ou TIFF. Os pesos moleculares dos fragmentos foram estimados usando-se o software GelAnalyzer (<http://www.gelanalyzer.com/>), com os seguintes parâmetros: *Rolling ball: 25, MW calibration: Log fit*. O programa compara a distância de migração das bandas no gel com as distâncias percorridas por fragmentos de tamanho conhecido presentes nos padrões de peso molecular. Bandas correspondentes a fragmentos menores do que 250 pb e maiores que 4300 pb não foram consideradas, baseado nas observações anteriores de que os valores extremos são mais propensos a erros no dimensionamento dos fragmentos durante o procedimento experimental (Coimbra, Grimont e Grimont, 1999).

Os perfis *cps*-RFLP experimentais assim obtidos foram primeiramente comparados com os perfis calculados *in silico*, posteriormente comparados entre si, usando-se o algoritmo de alinhamento do software MST.

3.3.2.6 Avaliação da reprodutibilidade dos perfis experimentais *cps*-RFLP

Experimentos usando réplicas técnicas foram realizados para avaliar a reprodutibilidade dos ensaios para obtenção dos perfis experimentais *cps*-RFLP e para a

determinação da variação intra- e inter-gel nos valores estimados dos pesos moleculares dos fragmentos de restrição obtidos para um mesmo isolado. Estes limites de variação experimental foram usados para verificar a calibração do algoritmo do *software* MST.

Os perfis experimentais *cps*-RFLP foram obtidos no mínimo duas vezes para um mesmo isolado para avaliar a reprodutibilidade da técnica, sendo que um total de 157 sorotipagens moleculares foram realizadas durante o período do estudo.

3.3.2.7 Separação dos fragmentos da digestão enzimática no instrumento Bioanalyzer

A separação dos fragmentos e a estimativa de seus pesos moleculares também foram realizadas em sistema Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). O Bioanalyzer é uma plataforma para quantificação, avaliação da qualidade e determinação dos tamanhos dos fragmentos de amostras de DNA, RNA ou proteína. O sistema faz uso de microcapilares para separação das amostras e, assim, reduz à quantidade de material necessário para cada corrida produzindo resultados em 30-40 minutos.

O Agilent DNA 7500 Kit (Agilent Technologies, Palo Alto, CA), específico para separação e quantificação de fragmentos de 100 a 7500 bp foi utilizado de acordo instruções do fabricante.

Os perfis *cps*-RFLP experimentais assim obtidos foram comparados com os perfis calculados *in silico* por meio do algoritmo de alinhamento do software MST.

3.4 Análise estatística

Foi utilizado o teste de *Wilcoxon rank-sum* (R, version 2.15.1) para comparação dos resultados obtidos com as diferentes enzimas de restrição. Valores de *p* inferiores a 0,05 foram considerados significativos.

3.5 Criação do banco de dados com perfis *cps*-RFLP

Os padrões *cps*-RFLP obtidos *in silico* foram carregados no banco de dados do software MST (Coimbra *et al.*, 2010), disponível em <http://www.cebio.org/mst>. Em adição, o banco de dados foi complementado com os padrões experimentais que não corresponderam aos *in silico*.

3.6 Requisitos legais

Autorização da Fundação Ezequiel Dias para utilização da coleção de *Streptococcus pneumoniae* na validação experimental da ferramenta de sorotipagem molecular (ANEXO VII).

4.1 Resultados das análises *in silico*

A análise com o software REMAP levou à seleção de cinco enzimas de restrição que apresentaram entre quatro e 26 sítios de clivagem em todos os clusters *cps* dos 90 sorotipos para os quais foi possível obter as sequências de DNA no GenBank. As enzimas selecionadas foram as seguintes: BslFI, Eco57MI, HindII, StyI, XhoII. Na análise subsequente, geramos os bancos de dados com os pesos moleculares dos fragmentos de restrição do cluster *cps* calculados para cada uma dessas cinco enzimas conforme descrito no item 4.3.1. A partir das matrizes de distância par-a-par de todos os sorotipos foram calculadas a mediana, a média e o desvio-padrão das distâncias entre os perfis, assim como o número de pares de sorotipos indistinguíveis para cada enzima. Nesta etapa, também geramos o dendograma de distância dos perfis *cps*-RFLP e a representação esquemática dos perfis de restrição do cluster *cps* para cada enzima selecionada (figuras 6, 7, 8, 9 e 10). Na tabela 2, encontram-se os resultados das análises estatísticas realizadas para identificação da enzima mais discriminante.

Tabela 2: Resultado das análises estatísticas realizadas para seleção da enzima de restrição mais discriminante.

Enzimas	Número de pares indistinguíveis	Mediana das distâncias entre todos os pares	Média das distâncias entre todos os pares ± Desvio Padrão	Wilcoxon rank-sum test em comparação com XhoII (Valor- <i>p</i>)
BslFI	10	29,6	31,96±14,98	$< 2,2e^{-16}$
Eco57MI	15	30,9	33,02±14,74	$7,6e^{-10}$
HindII	15	29,6	31,54±14,90	$< 2,2e^{-16}$
StyI	15	32,3	35,32±17,95	0,066
XhoII	7	32,4	36,06±17,93	-

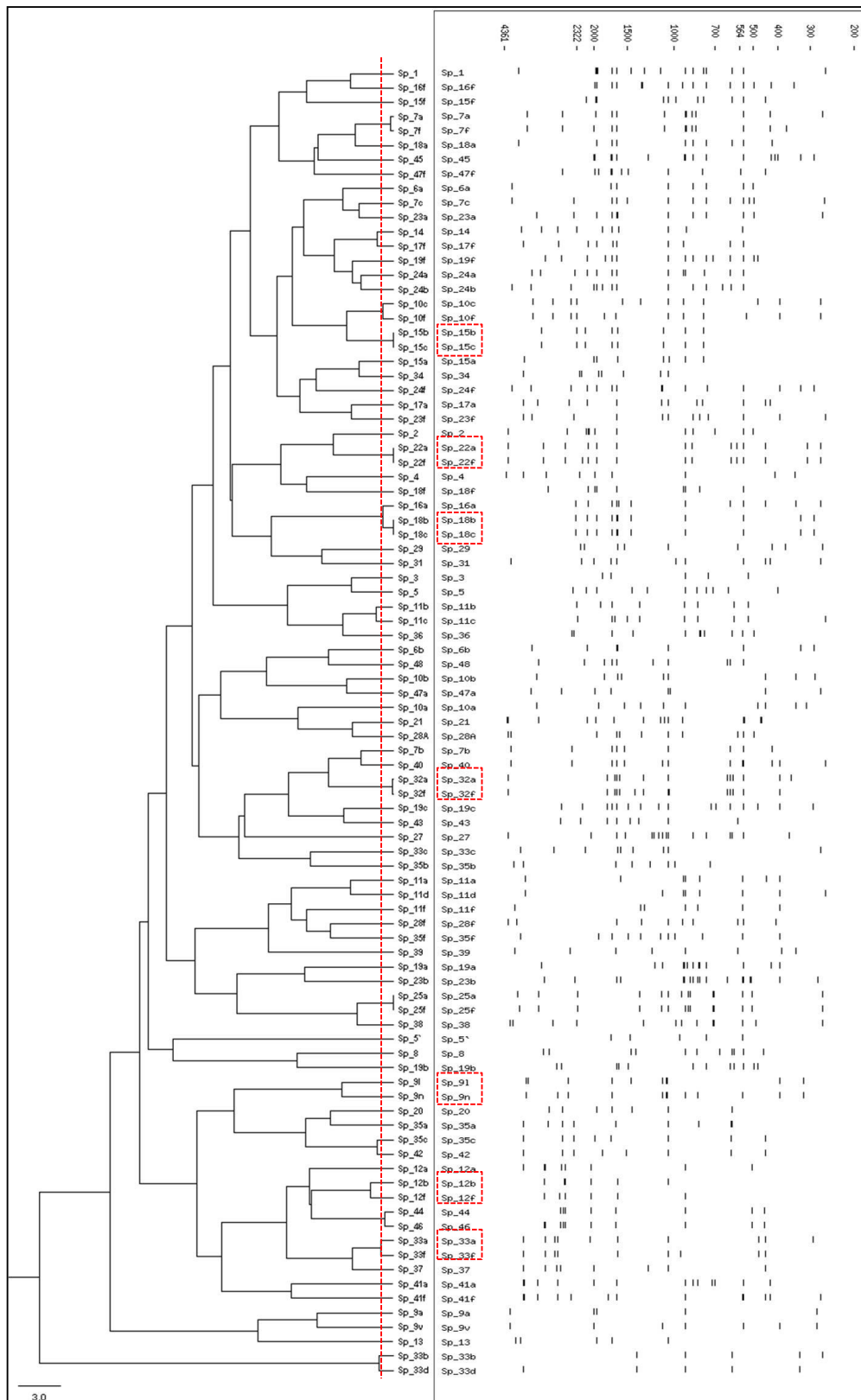


Figura 6: Representação esquemática do dendrograma de distância entre os perfis de restrição gerados pela enzima BslFI.

Painel direito, padrões *cps*-RFLP gerados pela digestão *in silico*. A linha tracejada representa o limiar do software MST (escore = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

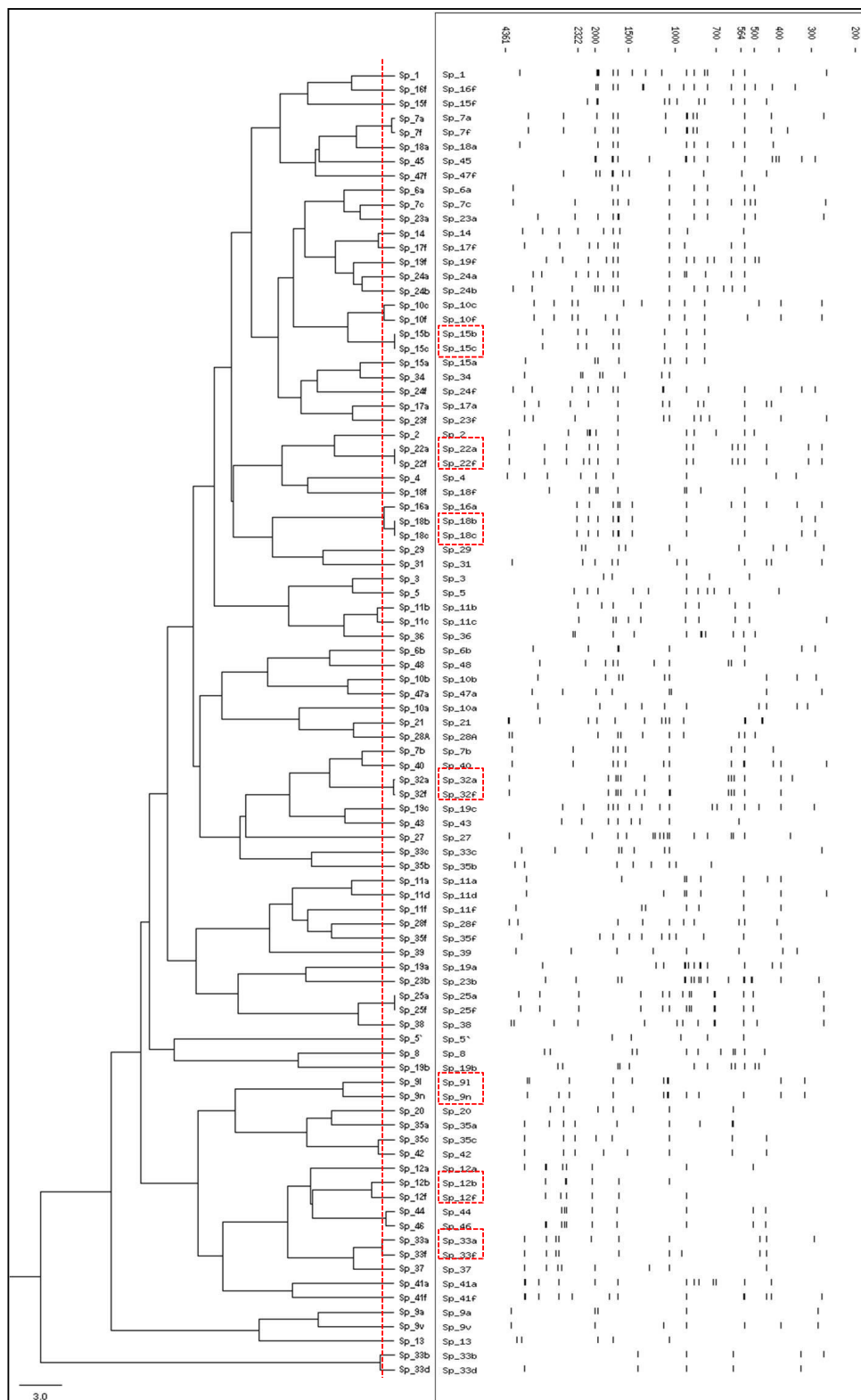


Figura 7: Representação esquemática do dendrograma de distância entre os perfis de restrição gerados pela enzima Eco57MI.

Painel direito, padrões *cps*-RFLP gerados pela digestão *in silico*. A linha tracejada representa o limiar do software MST (score = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

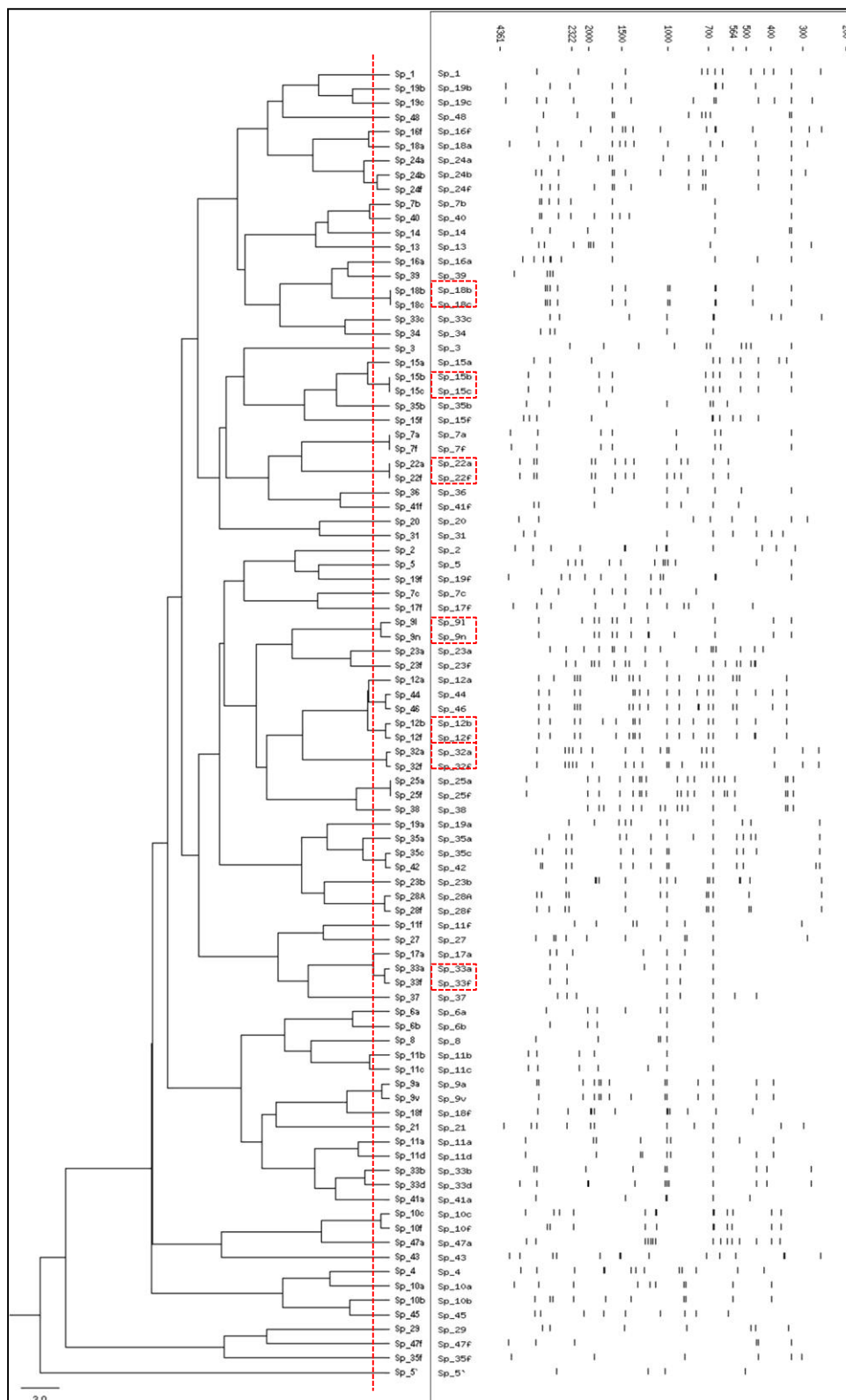


Figura 8: Representação esquemática do dendrograma de distância entre os perfis de restrição gerados pela enzima HindIII.

Painel direito, padrões *cps*-RFLP gerados pela digestão *in silico*. A linha tracejada representa o limiar do software MST (escore = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

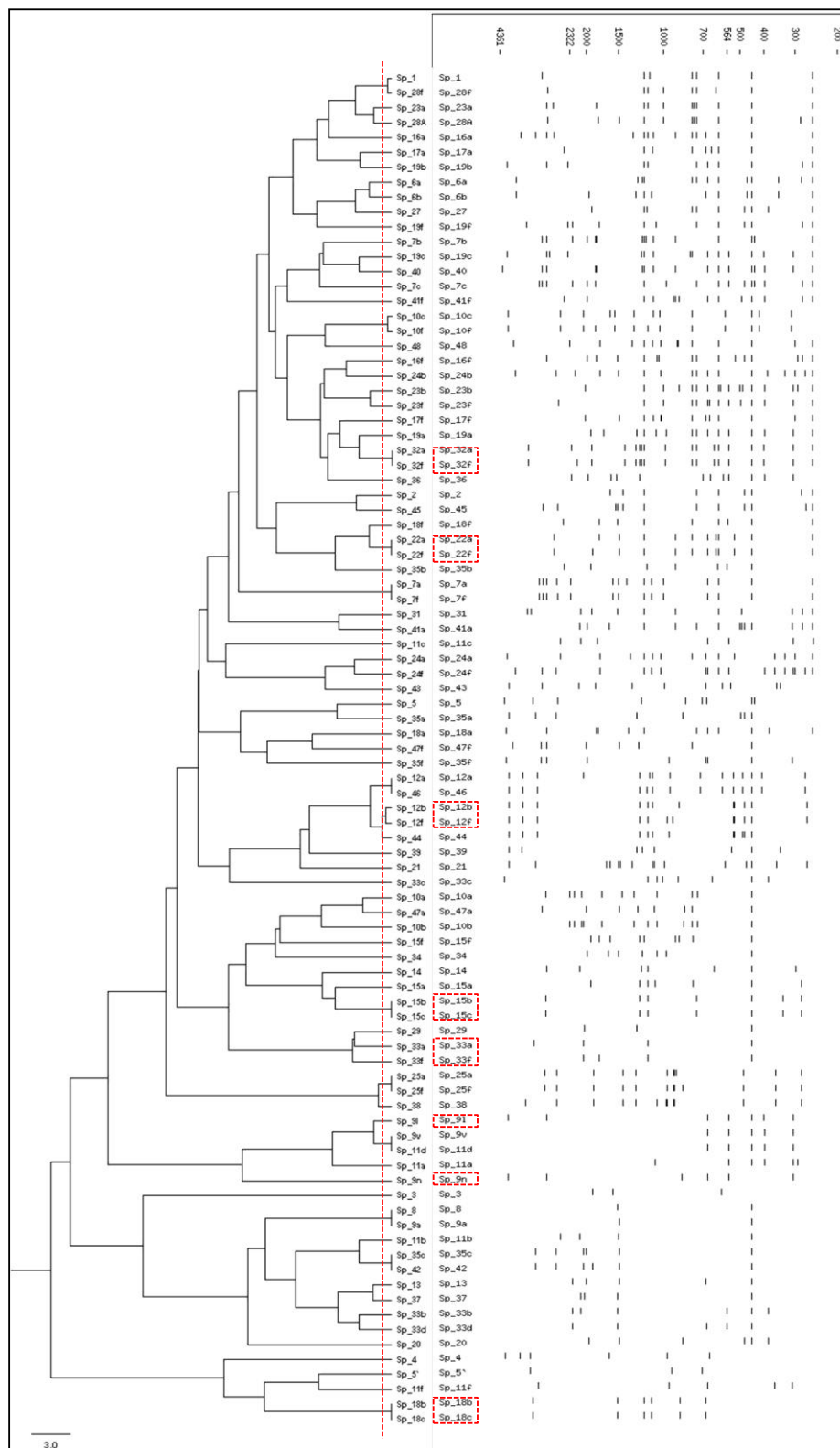


Figura 9: Representação esquemática do dendrograma de distância entre os perfis de restrição gerados pela enzima StyI.

Painel direito, padrões *cps*-RFLP gerados pela digestão *in silico*. A linha tracejada representa o limiar do software MST (escore = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

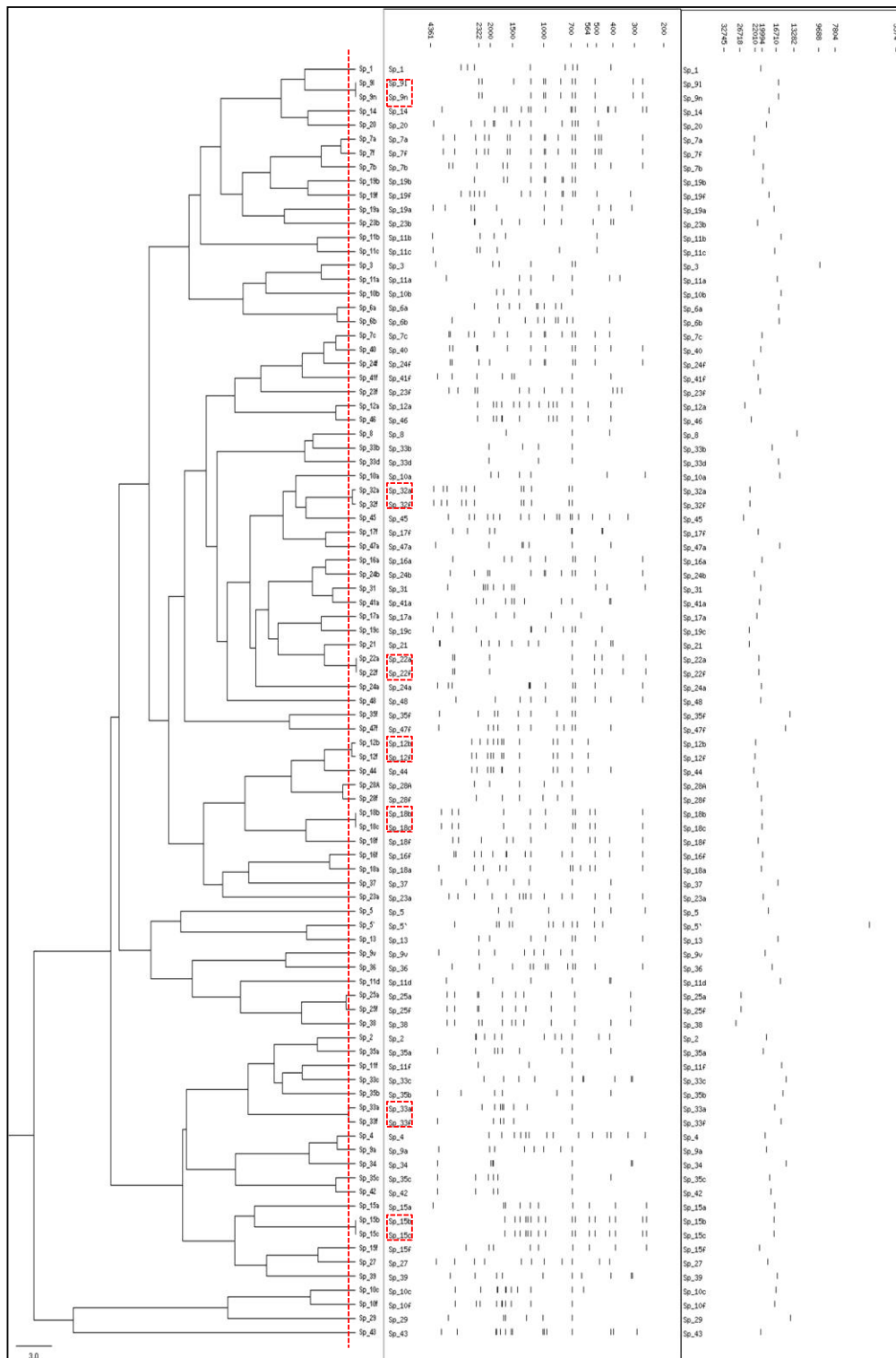


Figura 10: Representação esquemática do dendrograma das distâncias entre os perfis de restrição gerados pela enzima XhoII.

Painel central, padrões *cps*-RFLP gerados pela digestão *in silico*. Painel direito apresenta o tamanho do *locus cps* de cada sorotipo. A linha tracejada representa o limiar do software MST (score = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

A enzima XhoII foi selecionada na etapa de análises *in silico* por apresentar a maior média de distância entre perfis *cps*-RFLP gerados pelo software RESTRICT e o menor número de pares de perfis *cps*-RFLP indistinguíveis pelo software MST (Tabela 2). Esta enzima é uma endonuclease de restrição do tipo II que se liga e cliva a fita dupla de DNA com alta especificidade. Utiliza magnésio como cofator e produz fragmentos com extremidades 5' fosforiladas. A XhoII, identificada na bactéria *Xanthomonas holcicola*, reconhece sequências palindrômicas compostas por seis pares de bases, (A/G)[▼]GATC (T/C) / (T/C) CTAG[▲](A/G) (Chatterjee, 1994). XhoII cliva cada locus *cps* em pelo menos quatro e no máximo 26 posições diferentes. Os fragmentos de restrição menores que 250 pb e maiores que 4300 pb foram excluídos das análises com base em nossa experiência prévia (Coimbra, Grimont e Grimont, 1999; Coimbra *et al.*, 2000). A análise da faixa de pares de base mais discriminante confirmou a escolha da faixa descrita acima (Figura 12). Na análise final, os padrões *cps*-RFLP tinham três a 17 fragmentos que variavam entre 254 pb e 4.274 pb (Figura 10). A linha de corte de similaridade pré-estabelecida para o MST (escore MST = 1,5) foi validada como a mais discriminante entre todas as outras testadas (Figura 11). O MST distinguiu, sem ambiguidades, a quase totalidade dos padrões *cps*-RFLP sorotipo-específicos entre todos os outros no banco de dados. As sete exceções de pares de sorotipos que o programa não conseguiu distinguir com um limiar de escore MST de 1,5 foram 9L/9N, 12B/12F, 15B/15C, 18B/18C, 22A/22F, 32A/32F, 33A/33F (Figura 13). Os pares indistinguíveis eram do mesmo sorogrupo, o que pode ser explicado pelo fato destes serem imunologicamente e geneticamente similares. Ainda assim, para o conjunto de perfis *cps*-RFLP obtidos *in silico*, o software MST alcançou uma especificidade de 84,4% para sorotipagem e de 100% para sorogrupagem.

As análises *in silico* foram concluídas e os perfis *cps*-RFLP calculados para todos os 90 sorotipos de *S. pneumoniae* foram depositados no banco do software MST para *Streptococcus pneumoniae* (Anexo II).

Escore MST	Número de Pares Indistinguíveis	Pares Indistinguíveis													
		9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f							
1.5	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f							
1.6	8	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f						
2.2	9	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f					
2.5	10	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f				
3.1	12	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4		
3.3	14	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
3.4	15	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
3.8	16	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
3.9	17	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
4.1	18	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
4.2	19	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
4.4	20	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
4.9	21	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.1	24	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.2	25	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.3	26	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.4	27	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.5	29	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.7	30	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.8	33	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
5.9	34	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	
6.0	35	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	25a/25f	28a/28f	7a/7f	6a/6b	9a/4	*	

*Pares indistinguíveis não mostrados na figura.

Figura 11: Resultado da análise para escolha do escore MST mais distinguível.

Faixas analisadas	Mediana	Média	Desvio Padrão	Número de Pares Indistinguíveis (Score=1.5)	Pares Indistinguíveis												
					9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
200-20000	36,3	38,34353	17,84912	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
200-10000	36,3	38,20449	17,83145	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
200-7000	35,35	37,57339	17,90511	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
200-5000	34,1	36,60473	18,00675	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
250-4300	32,4	36,06	17,93	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
300-5000	33,3	35,64892	18,1677	7	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f						
400-5000	32,9	35,3376	18,19009	9	7a/7f	11a/3	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f				
500-5000	32,7	35,46636	18,99067	9	7a/7f	11a/3	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f				
500-7000	34	36,6193	18,93275	9	7a/7f	11a/3	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f				
700-5000	30,6	33,68531	18,89151	10	11a/3	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	31/41a	10a/48			
1000-5000	27,7	30,54176	17,72688	10	11a/3	9l/9n	12b/12f	15b/15c	18b/18c	22a/22f	32a/32f	33a/33f	35c/42	35c/42			
1500-5000	31,5	35,19723	22,76927	29	6a/6b	7a/7f	7f/8	9l/9n	12b/12f	15b/15c	1/18b	1/18c	18b/18c	22a/22f	*		

*Pares indistinguíveis não mostrados na figura.

Figura 12: Resultado das análises para escolher a faixa de peso molecular mais discriminante.

4.2 Validação da enzima de restrição selecionada frente à endonuclease HinfI

A enzima HinfI já tinha sido proposta para utilização na sorotipagem molecular de um conjunto de 46 sorotipos (Batt *et al.*, 2005). Foi gerado um banco de dados com os padrões *cps*-RFLP previstos por digestão *in silico* do *cluster cps* dos 90 sorotipos com a enzima HinfI. Os padrões *cps*-RFLP previstos para cada sorotipo foram comparados com o algoritmo de alinhamento de perfis de restrição do *software* MST e os escores dos alinhamentos par-a-par foram registrados. A tabela 3 apresenta os resultados das análises estatísticas da comparação entre as enzimas XhoII e HinfI. A HinfI produziu padrões *cps*-RFLP que tinham de nove a 44 fragmentos que variavam entre 250 pb e 4.120 pb (Figura 14). A comparação dos dendogramas de distância e dos perfis de restrição das duas enzimas evidenciou que a enzima XhoII diferencia melhor entre os perfis *cps*-RFLP e gera um menor número de fragmentos mais dispersos entre os limiares de 250 e 4300 pb. A especificidade calculada para a endonuclease HinfI foi de 42,2% para sorotipagem e 57,7% para sorogrupagem, resultados bem inferiores a especificidade da XhoII (84,4% e 100%, respectivamente). Os resultados das análises estatísticas apresentados na tabela 3 confirmaram o maior poder discriminatório da enzima XhoII.

Tabela 3: Resultado da análise estatística para comparar as enzimas XhoII e HinfI.

Enzimas	Número de pares indistinguíveis	Mediana das distâncias entre todos os pares	Média das distâncias entre todos os pares ±Desvio Padrão	Wilcoxon rank-sum test frente XhoII (Valor-p)
HinfI	34	8,8	10,56±6,65	$< 2.2e^{-16}$
XhoII	7	32,4	36,06±17,93	-

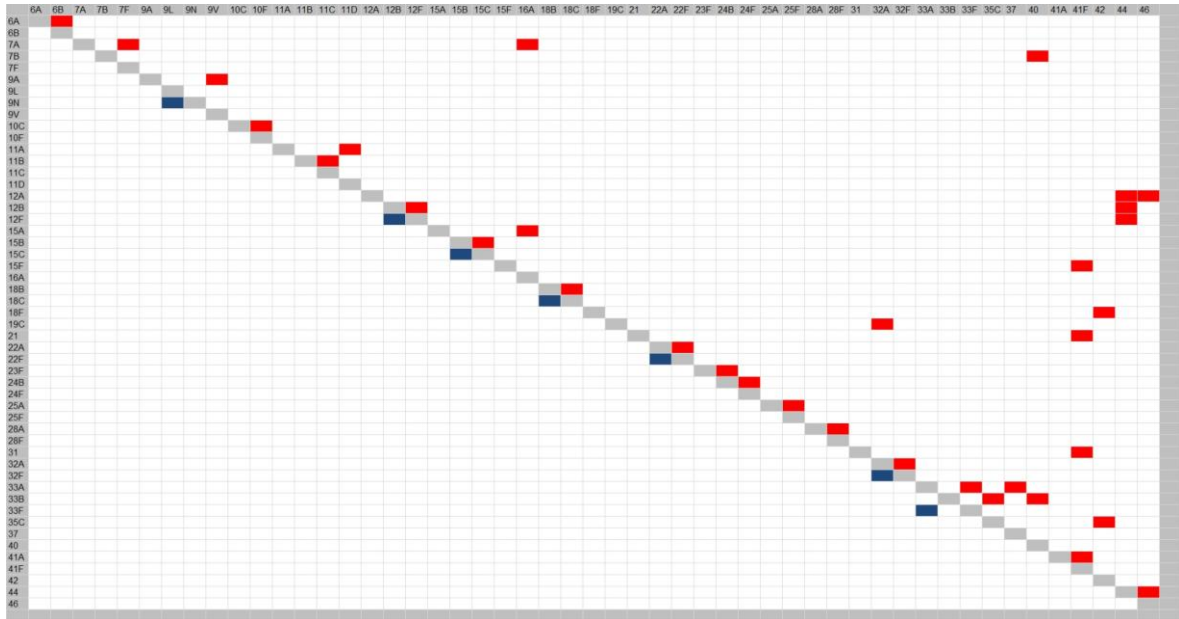


Figura 13: A substituição da endonuclease HinfI por XhoII aumentou a especificidade do *cps*-RFLP.

Caixas azuis = *cps*-RFLP usando XhoII gera menor número de pares indistinguíveis (7).
 Caixas vermelhas = usando HinfI, 34 pares de padrões *cps*-RFLP são indistinguíveis pelo MST. Especificidade: XhoII = 84,4%; HinfI = 42,2%.

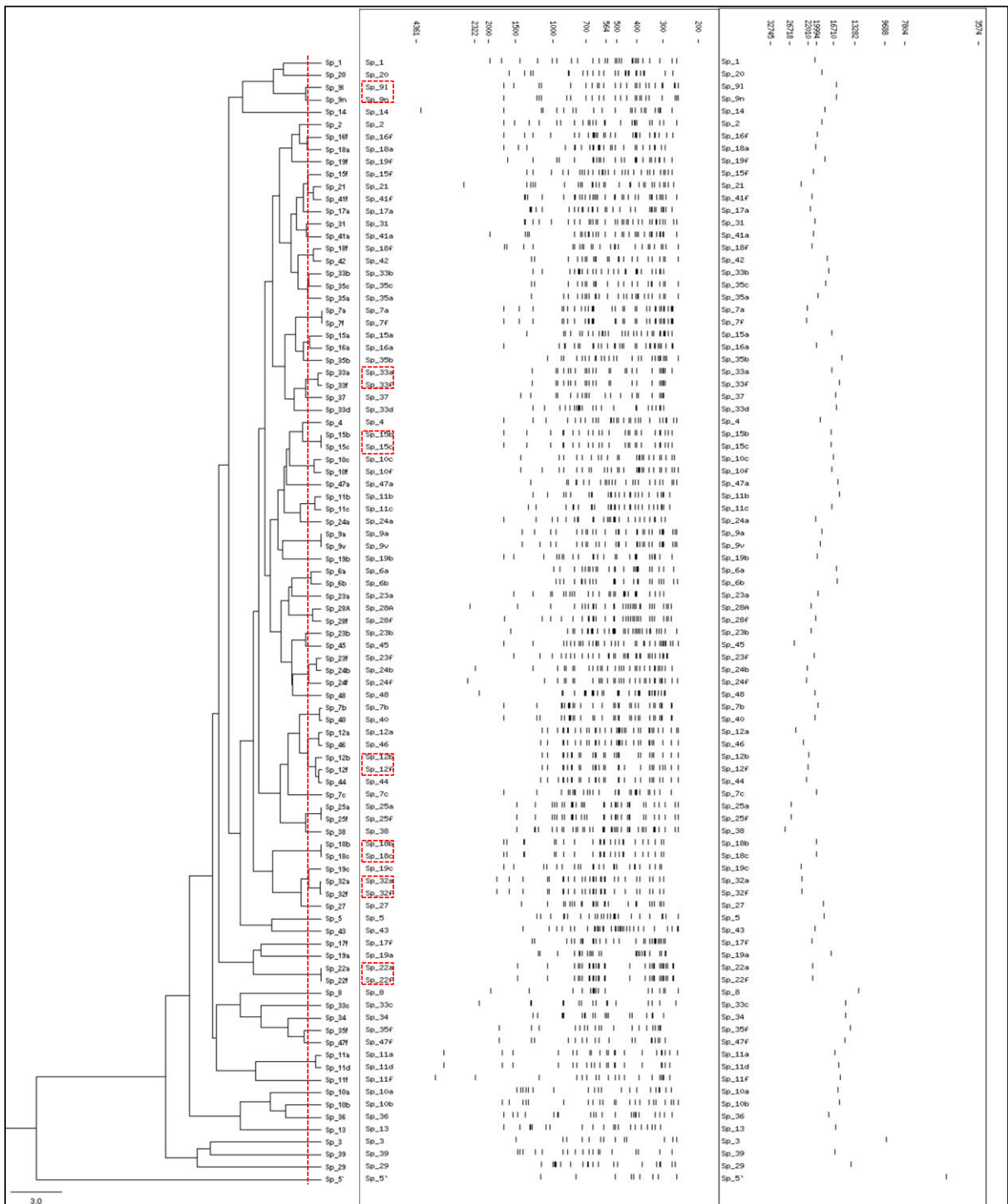


Figura 14: Representação esquemática do dendrograma de distância entre os perfis de restrição gerados pela enzima HinfI.

Painel central, padrões *cps*-RFLP gerados pela digestão *in silico*. Painel direito apresenta o tamanho do *locus cps* de cada sorotipo. A linha tracejada representa o limiar do software MST (escore = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

4.3 Validação experimental

4.3.1 Amplificação e restrição do cluster *cps*

O cluster *cps* de um conjunto de 46 cepas de *Streptococcus pneumoniae* epidemiologicamente não relacionadas, representando 32 sorotipos patogênicos prevalentes no estado de Minas Gerais foi amplificado como descrito na metodologia (Figuras 15 a 20). Para nosso espanto, não tivemos sucesso na amplificação das regiões *cps* de dois isolados do sorotipo 3 e de um do sorotipo 22F. Isto é surpreendente, uma vez que Batt e colaboradores (2005), amplificaram com êxito a região *cps* destes dois sorotipos usando os mesmos iniciadores que utilizamos.

Os produtos da amplificação variavam entre 17.142 e 26.106 pares de bases (Anexo IV). Os perfis de restrição *cps*-RFLP, claramente identificáveis e reprodutíveis, eram compostos por quatro a 17 fragmentos, após separação dos fragmentos em gel de agarose (Figura 22). Os experimentos usando réplicas técnicas, realizados para determinação da variação intra- e inter-gel nos valores estimados de peso molecular dos fragmentos de restrição obtidos para um mesmo isolado, demonstraram uma variação inter-gel máxima de 4,33% na faixa inferior do gel (250 a 500 pb) e de 2,23% na faixa superior do gel (500 a 4300 pb). A variação intra-gel foi inferior à inter-gel (1,99% na faixa inferior e 1,74% na faixa superior do gel). Estes limites foram inferiores aos valores *default* do software MST (limite inferior de 7% e na faixa superior de 3,5%) e, por isso os valores *default* foram mantidos.

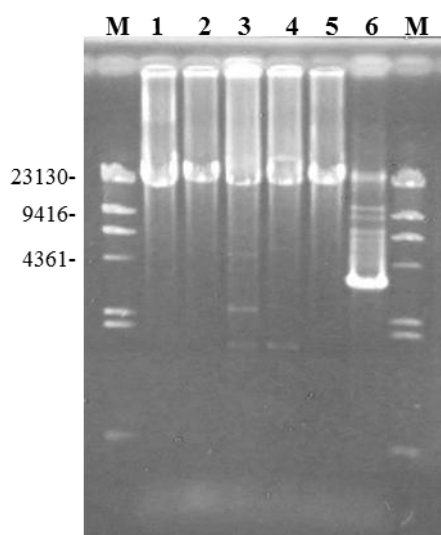


Figura 15: Produto da amplificação do cluster *cps* dos sorotipos 19F, 18C, 19A, 6B, 18B, 14.

Canaletas: M = marcador de peso molecular (*Lambda DNA HindIII*), 1 = 24/12 (19F), 2 = 120/11 (18C), 3 = 124/11 (19A), 4 = 143/11 (6B), 5 = 240/11 (18B), 6 = 328/11 (14).

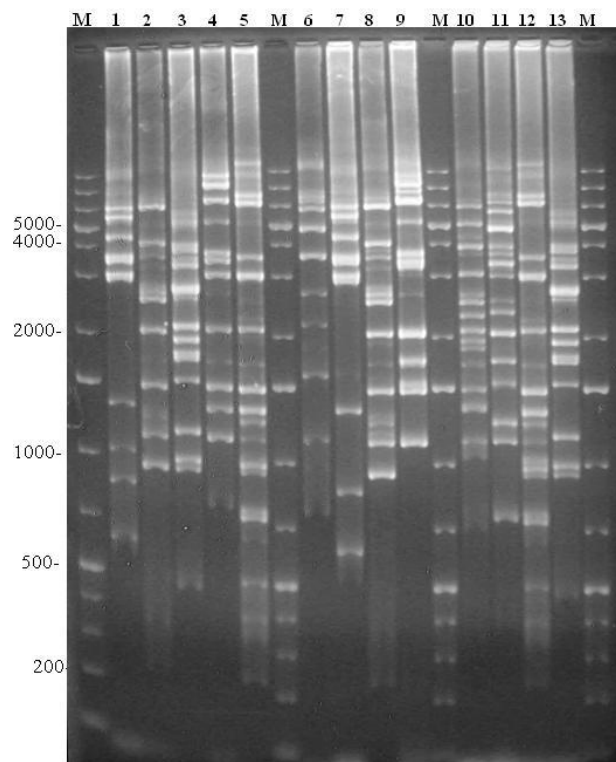


Figura 16: Perfis de restrição com a enzima XhoII *cps*-RFLP dos sorotipos 29, 9N, 19F, 6C, 19A, 29, 6A, 19F, 18B, 6B.

Canaletas: M = marcador de peso molecular (*GeneRuler 1 kb Plus DNA Ladder*), 1 = 79/11 (29), 2 = 103/11 (6A), 3 = 149/11 (9N), 4 = 24/12 (19F), 5 = 387/11 (6C), 6 = 124/11 (19A), 7 = 79/11 (29), 8 = 103/11 (6A), 9 = ATCC49619 (19F), 10 = 240/11 (18B), 11 = 143/11 (6B), 12 = 387/11 (6C), 13 = 149/11 (9N).

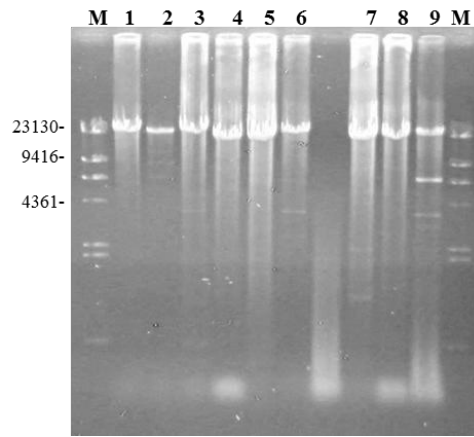


Figura 17: Produto da amplificação do cluster *cps* dos sorotipos 24F, 12F, 35B, 5, 10A, 9V, 9N e 7F.

Canaletas: M = marcador de peso molecular (*lambda Hind III DNA Ladder*), 1 = 80/11 (24F), 2 = 84/11 (12F), 3 = 120/11 (18C), 4 = 149/11 (9N), 5 = 192/11 (35B), 6 = 197/11 (5), 7 = 490/11 (10A), 8 = 511/11 (9V), 9 = 731/09 (7F).

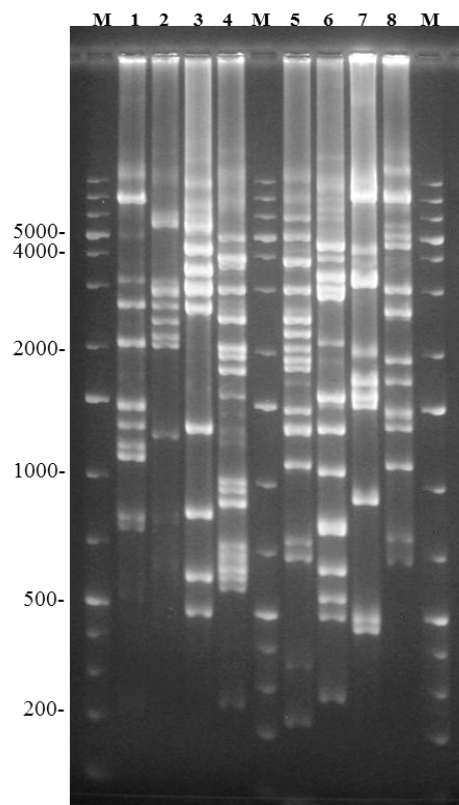


Figura 18: Perfis de restrição com a enzima XhoII *cps*-RFLP dos sorotipos 24F, 12F, 35B, 5, 18C, 10A, 9V e 7F.

Canaletas: M = marcador de peso molecular (*GeneRuler 1 kb Plus DNA Ladder*), 1 = 80/11 (24F), 2 = 84/11 (12F), 3 = 192/11 (35B), 4 = 197/11 (5), 5 = 120/11 (18C), 6 = 490/11 (10A), 7 = 511/11 (9V), 8 = 731/09 (7F).

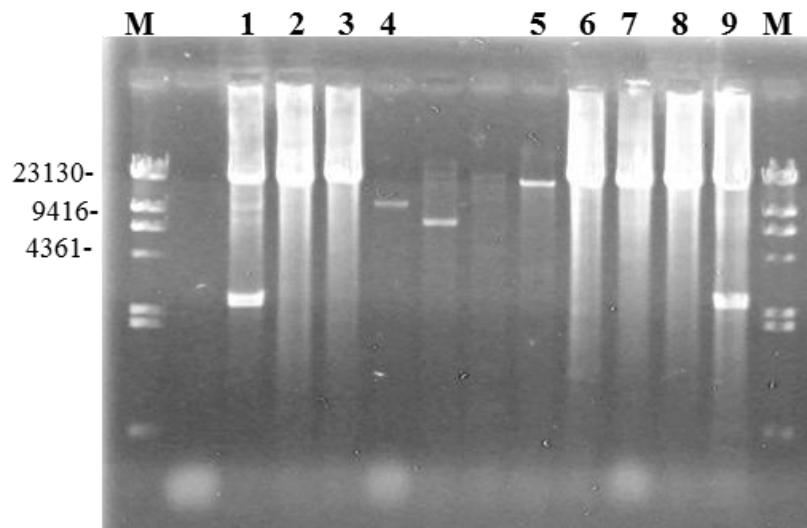


Figura 19: Produto da amplificação do cluster *cps* dos sorotipos 6C, 35B, 9N, 3, 6C, 10A, 6A/C, 9N e 6A/C.

Canaletas: M = marcador de peso molecular (*lambda Hind III DNA Ladder*), 1 = 127/10 (6C), 2 = 144/09 (35B), 3 = 149/11 (9N), 4 = 159/11 (3), 5 = 387/11 (6C), 6 = 421/10 (10A), 7 = 619/08 (6A/C), 8 = 620/09 (9N), 9 = 1180/07 (6A/C).

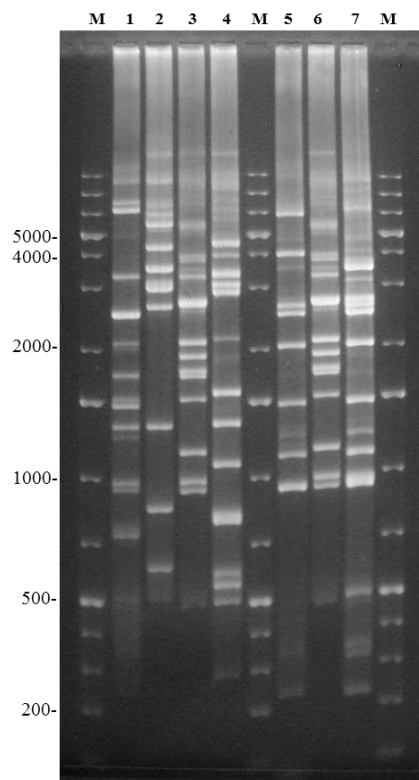


Figura 20: Perfis de restrição com a enzima XhoII *cps*-RFLP dos sorotipos 6C, 35B, 9N, 10A, 6A/C, 9N e 6A/C.

Canaletas: M = marcador de peso molecular (*GeneRuler 1 kb Plus DNA Ladder*), 1 = 127/10 (6C), 2 = 144/09 (35B), 3 = 149/11 (9N), 4 = 421/10 (10A), 5 = 619/08 (6A/C), 6 = 620/09 (9N), 7 = 1180/07 (6A/C).

Os quarenta e seis perfis *cps*-RFLP experimentais foram testados contra o banco de dados virtual gerado na etapa de análise *in silico* utilizando o software MST (Figura 21). Os escores do grupo experimental variaram de zero a 17,5. Nesta etapa também geramos o dendograma de distância entre os perfis experimentais *cps*-RFLP e a representação esquemática dos perfis de restrição do cluster *cps* (Figura 22).

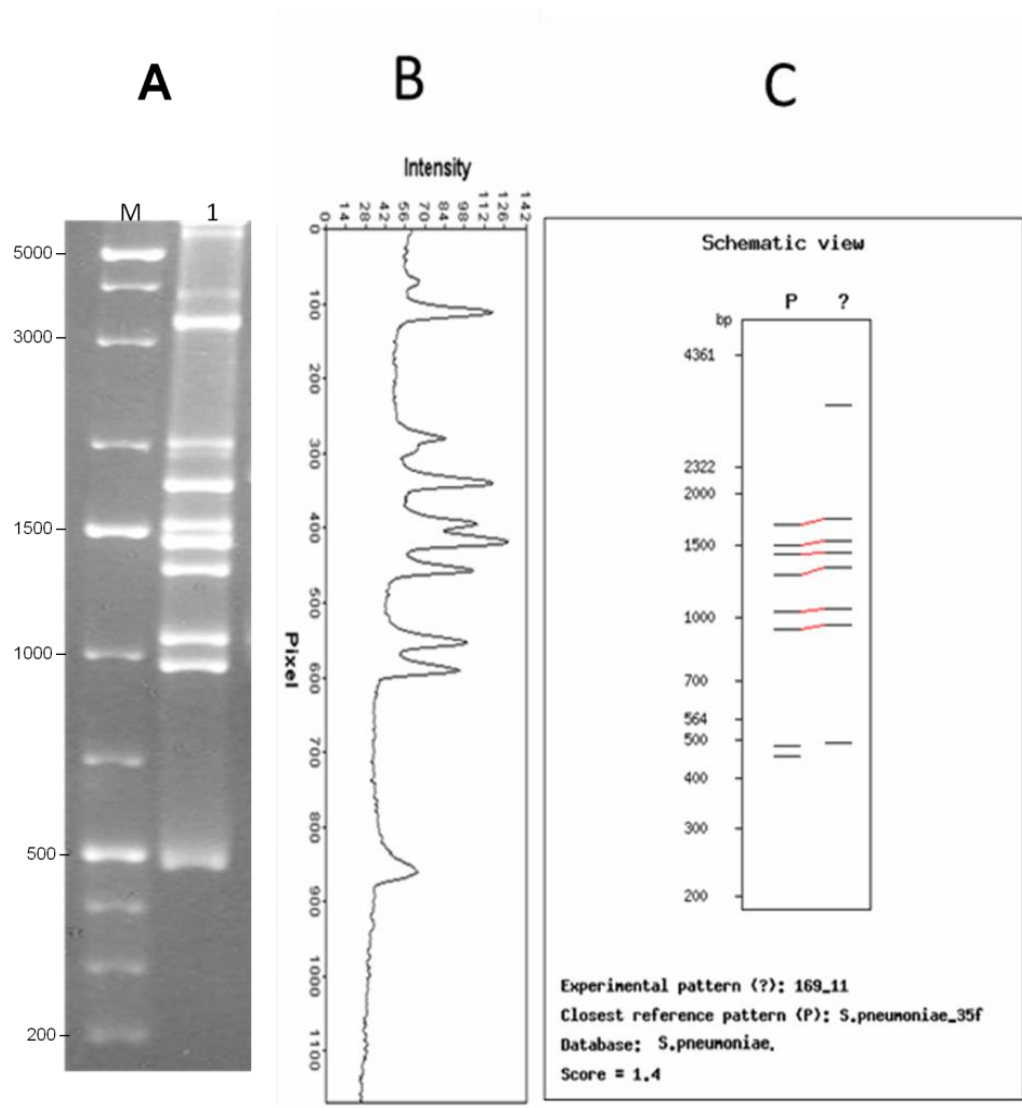


Figura 21: O perfil *cps*-RFLP experimental obtido demonstra estreita similaridade com aquele previsto *in silico*.

A) Canaletas: M = marcador de peso molecular; 1 = perfil *cps*-RFLP do sorotipo 35F; B) perfil *cps*-RFLP do sorotipo 35F (isolado 169/11) analisado com o software GelAnalyzer; C) *output* do software MST, mostrando a representação esquemática do perfil experimental *cps*-RFLP do *S. pneumoniae*, sorotipo 35F (isolado 169/11) alinhado com o padrão de referência previsto *in silico*.

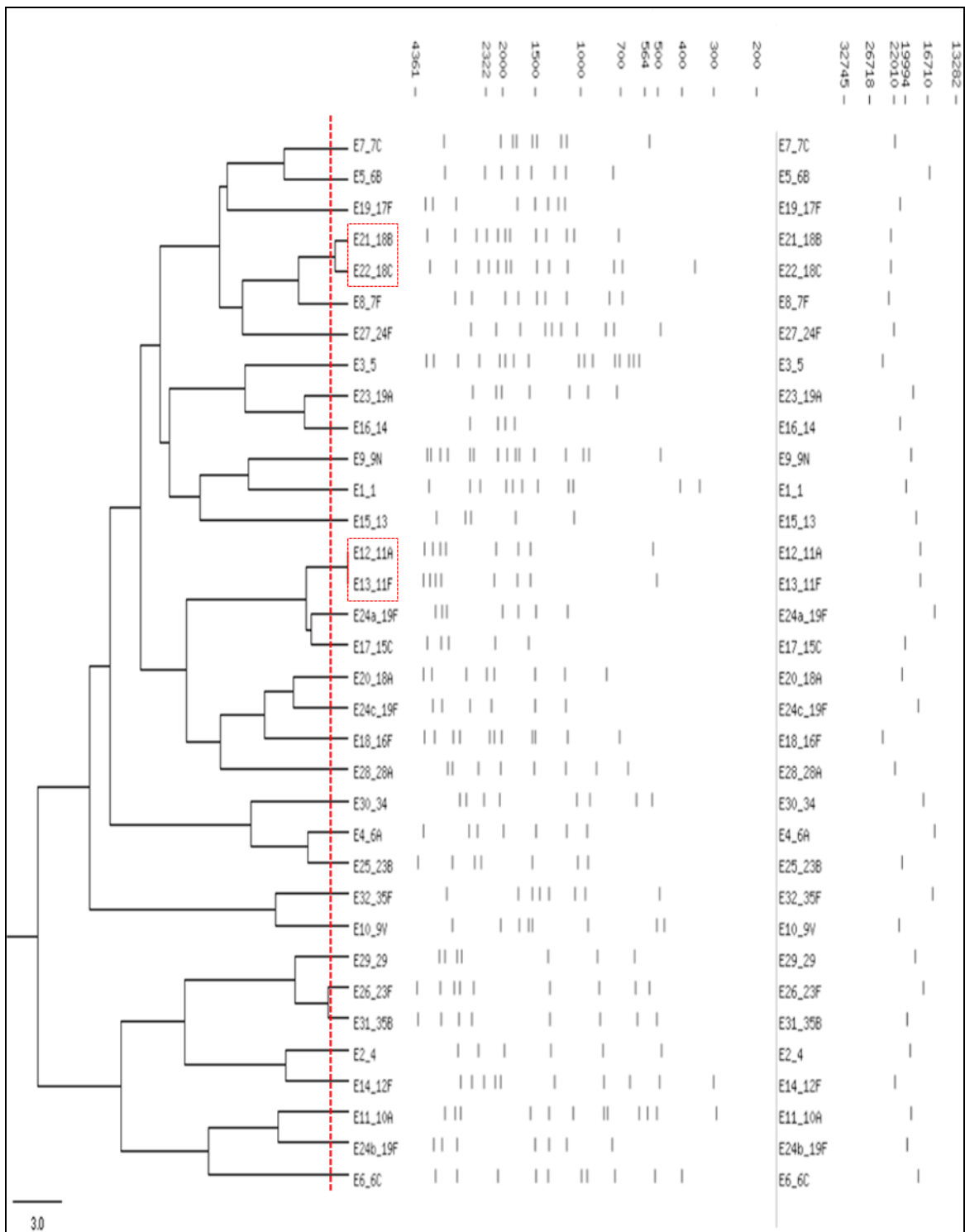


Figura 22: Representação esquemática do dendrograma de distância entre os perfis de restrição experimentais gerados pela enzima XhoII.

No painel central os “fingerprints” gerados pela digestão com XhoII (padrões experimentais *cps*-RFLP) e o painel direito apresenta o tamanho total da região *cps* de cada sorotipo. A linha tracejada representa o limiar do software MST (score = 1,5) e a caixa vermelha identifica os pares de sorotipos indistinguíveis usando XhoII.

Os resultados da comparação dos perfis experimentais com o banco de dados *in silico* utilizando o algoritmo de alinhamento do MST estão compilados na tabela 4.

Tabela 4: Resultado das análises dos perfis experimentais contra o banco de dados virtual gerado na etapa de análise *in silico* utilizando o software MST.

Escore MST	Identificação em nível de sorotipo	Identificação em nível de sorogrupo	Não identificou o sorotipo ou sorogrupo no banco de dados <i>in silico</i>
≤ 1,5 ^a	1, 6A, 12F, 35F, 14, 9V, 19A (n=7; 21,87%)	7F (7A) ^b (n=1; 3,12%)	-
>1,5	4, 5, 7C, 10A, 34, 16F, 17F, 18A, 19F, 23B, 24F, 28A, 35B (n=13; 40,63%)	11A (11D), 11F (11D), 9N (9L), 18B (18F), 18C (18F) (n=5; 15,63%)	6C, 15C (22A), 13 (4), 23F (35B), 6B (11D), 29 (42) (n=6; 18,75%)
Total	20 (62,5%)	6 (18,75%)	6 (18,75%)

^aLinha de corte do escore de similaridade pré-estabelecida para o software MST para pneumococos. A identificação em nível de sorotipo ocorre quando o escore de alinhamento entre os dois perfis *cps*-RFLP é menor ou igual a 1,5.

^bEntre parênteses estão representados os sorotipos dos perfis *in silico* retornados pelo software.

O perfil do sorotipo 6C não tinha correspondente no banco de dados de perfis preditos, pois a sequência completa de DNA do seu *cps* até o presente momento ainda não tinha sido publicada. Os resultados completos da comparação entre os perfis *in silico* e experimentais estão descritos no Anexo V.

Não foi possível amplificar o *cps* do sorotipo 22F após sucessivas tentativas e o rendimento da reação de amplificação do sorotipo 3 não foi suficiente para gerar o seu perfil *cps*-RFLP.

Na análise final da ferramenta de sorotipagem molecular (*cps*-RFLP), os padrões *cps*-RFLP previstos e experimentais para cada sorotipo analisado foram comparados entre si com o algoritmo de alinhamento de perfis de restrição do *software* MST. A matriz de distância par-a-par entre todos os perfis *in silico* e experimentais demonstrou que o *software* MST não conseguiu distinguir nove pares de sorotipos para um escore MST de 1,5. Os pares indistinguíveis foram os seguintes 9L/9N, 12B/12F, 15B/15C, 18B/18C, 22A/22F, 32A/32F, 33A/33F, 7F/7A, 11A/11D. A inclusão de todos os perfis experimentais obtidos reduziria a especificidade da ferramenta de 84,4% para 80,2%, pois dois novos pares de perfis indistinguíveis (7F/7A, 11A/11D) seriam introduzidos a ferramenta (Figura 23).

As análises experimentais foram concluídas e os perfis *cps*-RFLP experimentais sem correspondentes no banco de dados *in silico* para um escore MST de 1,5 foram depositados no banco do *software* MST para *Streptococcus pneumoniae* (Anexo III).

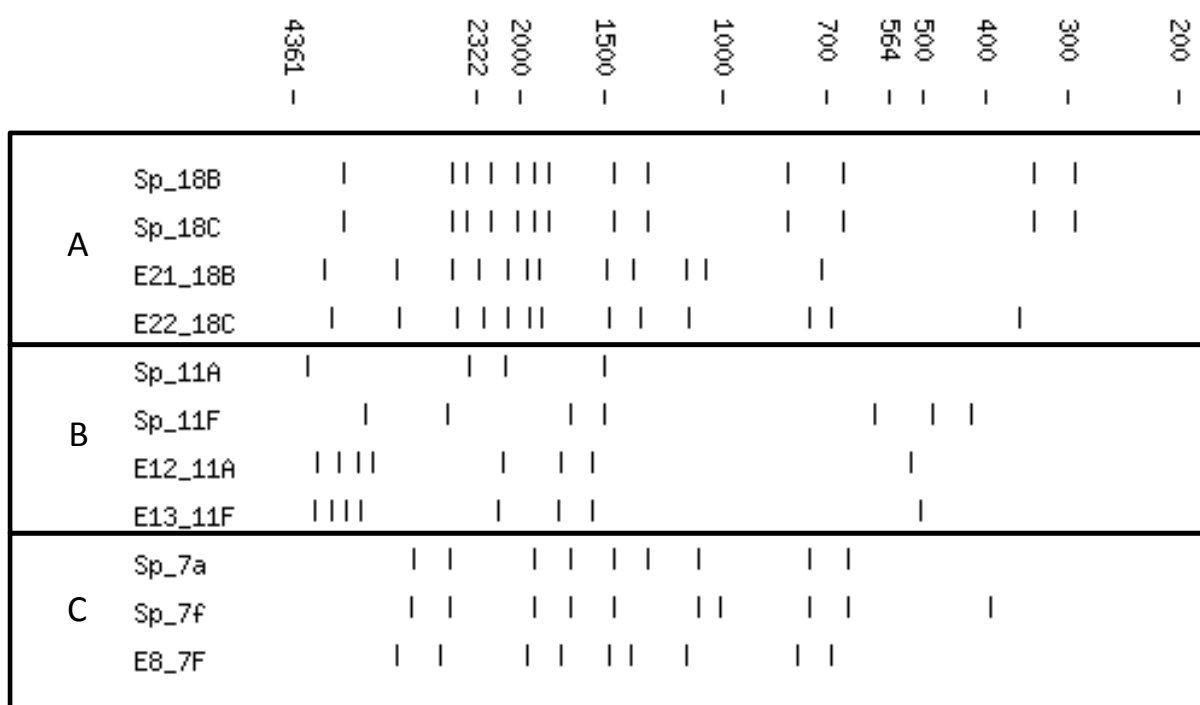


Figura 23: Representação esquemática dos dois pares de perfis experimentais indistinguíveis e do par *in silico*/experimental indistinguível.

A) Sp_18B/18C *in silico*, E21-E22_18B/18C experimentais; B) Sp_11A/11F *in silico*, E12-E13_11A/11F experimentais 11A/11F; C) Sp_7A e 7F *in silico*, E8-7F experimental.

4.3.2 A separação dos fragmentos e a estimativa de seus pesos moleculares usando o sistema Agilent 2100 Bioanalyzer

Um total de 63 produtos de restrição de *clusters cps* com XhoII foram analisadas com o sistema Agilent 2100 Bioanalyzer utilizando para esta finalidade o Agilent DNA 7500 Kit (Agilent Technologies, Palo Alto, CA), específico para separação e quantificação de fragmentos de 100 a 7500 bp (Figura 24).

A expectativa de aumentar a precisão na determinação do número de pares de base não se confirmou. Ficou evidente que na faixa do gel onde os fragmentos de restrição ficam mais concentrados (2000 a 4300 pb) o kit utilizado não produziu uma separação adequada dos fragmentos e também não foi capaz de determinar com exatidão e reprodutibilidade o tamanho estimado dos fragmentos de restrição, quando comparado aos perfis gerados nas corridas em gel de agarose.

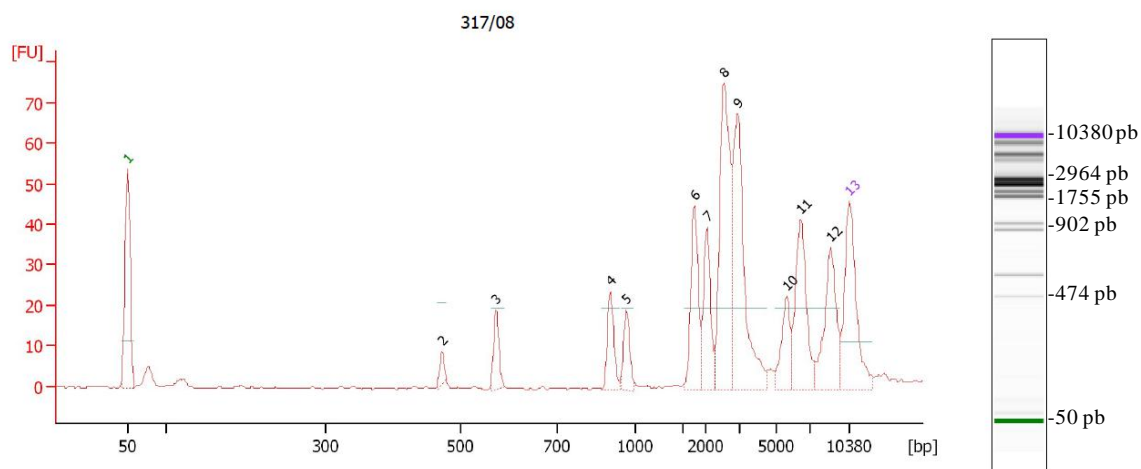
Sessenta e três perfis *cps*-RFLP experimentais obtidos no Agilent 2100 Bioanalyzer representando 30 sorotipos distintos foram testados contra o banco de dados virtual gerado na etapa de análise *in silico* utilizando o software MST. O software identificou o sorotipo correto, independente do escore MST, de 17 perfis (26,9%), sendo que em apenas um deles o escore foi menor ou igual a 1,5 e em dois perfis o escore foi $\leq 3,0$. Em três perfis o MST identificou um sorotipo diferente, mas dentro do mesmo sorogrupo (4,8%), e para os outros 42 perfis, o software não identificou seu correspondente no banco (66,6%). Os perfis obtidos para o sorotipo 6C não têm correspondente no banco de dados *in silico* do MST, conforme mencionado anteriormente. Os resultados da comparação estão descritos no Anexo VI. Os escores do grupo experimental no Agilent 2100 Bioanalyzer variaram de 1,5 a 20,7.

Os resultados acima evidenciaram que os perfis *cps*-RFLP obtidos no sistema Agilent 2100 Bioanalyzer eram mais distantes dos perfis *in silico* que aqueles perfis obtidos em géis de agarose. Por este motivo, somente os perfis *cps*-RFLP obtidos em géis de agarose foram carregados no banco de dados do software MST.

Assay Class: DNA 7500
 Data Path: \20...-06\2100 expert_DNA 7500_DE72904746_2013-09-06_20-44-00.xad

Created: 9/6/2013 8:44:00
 Modified: 9/6/2013 10:01:00

Electropherogram Summary Continued ...



Overall Results for sample 10 : 317/08

Peak table for sample 10 : 317/08

Peak	Size [bp]	Conc. [ng/μl]	Molarity [nmol/l]	Observations
1	50	8.30	251.5	Lower Marker
2	474	0.37	1.2	Number of peaks fou
3	574	1.03	2.7	
4	902	1.35	2.3	
5	965	1.02	1.6	
6	1,755	2.65	2.3	
7	2,042	2.15	1.6	
8	2,564	5.64	3.3	
9	2,964	5.89	3.0	
10	5,672	1.49	0.4	
11	6,471	3.22	0.8	
12	8,707	3.04	0.5	
13	10,380	4.20	0.6	Upper Marker

Figura 24: Representação esquemática do eletroferograma do perfil de restrição experimental gerado pela enzima XhoII da amostra 317/08 (sorotipo 34) .

4.4 Trabalhos apresentados em congressos científicos

Os resultados das análises *in silico* foram apresentados no XXI Congresso Latinoamericano de Microbiologia (XXI ALAM) que foi realizado no período de 28 de outubro a 1 de novembro de 2012, no Mendes *Convention Center*, na cidade de Santos, SP, Brasil (Anexo VIII) e os resultados das análises experimentais foram apresentados no X-meeting 2013 (*The Brazilian Symposium on Bioinformatic*), no formato de apresentação oral, realizado no período de 3 a 6 de novembro de 2013, na cidade de Recife, Pernambuco, Brasil (Anexo IX).

Apresentamos aqui um novo conjunto de ferramentas para sorotipagem molecular de *Streptococcus pneumoniae* (*cps*-RFLP) baseado no método proposto anteriormente para *Escherichia coli* (Coimbra *et al.*, 2000), que inclui: 1) um método de biologia molecular que produz perfis sorotipo-específico; 2) um banco de dados para armazenar os perfis de referência; e 3) um software para predizer o sorotipo da amostra clínica por comparação do seu perfil com aqueles no banco de dados.

Esta ferramenta de sorotipagem molecular denominada *cps*-RFLP consiste na amplificação da região *cps* que codifica os genes chaves para síntese do antígeno capsular. Os genes *dexB* e *aliA*, flanqueiam o *locus cps*, sendo altamente conservados entre todos os sorotipos de pneumococos. Por isso, foram usados como alvos de iniciadores para a amplificação do *locus cps* por PCR neste método. Esses genes, *dexB* e *aliA* não participam da síntese do CPS. O produto da amplificação é subsequentemente digerido com a enzima de restrição XhoII gerando perfis únicos (“*fingerprints*”) à eletroforese em gel de agarose, semelhantes a um código de barras, e permitindo, assim, a diferenciação de cada sorotipo. Os perfis de isolados clínicos, assim obtidos, são posteriormente analisados no software GelAnalyzer para calcular os pesos moleculares das bandas do perfil. Após esta etapa o perfil é carregado no software MST que compara os perfis dos isolados clínicos contra todos os perfis de referência contidos em seu banco de dados, retornando uma representação esquemática do alinhamento do perfil desconhecido contra o perfil do sorotipo mais semelhante do banco (score MST ≤ 1.5).

As análises *in silico* foram importantes, principalmente por dois aspectos. O primeiro foi por permitir a escolha da enzima de restrição mais discriminante, mesmo quando comparada a enzima previamente proposta para sorotipagem molecular de pneumococos (Batt *et al.*, 2005). O segundo foi por possibilitar a construção do banco de dados de perfis *in silico*, proporcionando assim, uma comparação entre os perfis construídos a partir de sequências de DNA do cluster *cps* de 90 sorotipos da coleção de cepas do mundo da Organização Mundial da Saúde, centro colaborador para referência e pesquisa em pneumococos, Statens Serum Institut, contra perfis experimentais obtidos de cepas isoladas entre os anos de 2007 e 2012, numa região geográfica bem definida (Minas Gerais – Brasil).

Trinta e quatro perfis *cps*-RFLP sorotipo-específicos, representando 32 sorotipos, foram testados contra o banco de dados obtidos *in silico*. As amostras clínicas de sete sorotipos foram corretamente identificadas pelo MST tendo apresentado escores de alinhamento com os respectivos perfis *in silico* inferiores a 1,5. O MST identificou matches para os perfis *cps*-RFLP de amostras de outros 13 sorotipos, mas com escores de alinhamento acima de 1,5. Os perfis *cps*-RFLP experimentais de amostras de outros cinco sorotipos

apresentaram alinhamentos com escore acima de 1,5 com perfis *in silico* de sorotipos diferentes, mas dentro do mesmo sorogrupo (Tabela 4). Apenas o par indistinguível 7A/7F obteve escore MST igual à zero. Esta cepa será novamente sorotipada pela reação de Quellung e a inclusão de seu perfil no banco de dados do software MST está condicionada a este resultado. Mavroidi e colaboradores (Mavroidi *et al.*, 2007) relataram alta similaridade entre as sequências do cluster *cps* do par 7A/7F, e que as estruturas do CPS destes dois sorotipos são muito semelhantes. Ao tipo 7A falta somente o ramo lateral D -Galp-(β 1-2)- α - D -Galp presente no tipo 7F, e esta diferença pode ser atribuída por uma mutação *frameshift* no gene *wcwD*.

O método *cps*-RFLP foi mais eficiente quando os fragmentos de restrição foram separados em géis de agarose. O sistema Agilent 2100 Bioanalyzer quando comparado com a eletroforese em géis de agarose, demonstrou menor poder de separação dos fragmentos na faixa de 2000 a 4000 pb, justamente a região mais informativa (Figura 24).

A sensibilidade do método molecular foi de 94,1%, (porque o método não foi capaz de tipar as cepas dos sorotipos 22F e 3 dos 34 sorotipos analisados, 32 sorotipos geraram perfis experimentais *cps*-RFLP). No caso específico do sorotipo 3, cuja incidência aumentou do ano de 2011 para 2012, conseguimos amplificar o seu cluster *cps*, mas o rendimento da reação não foi satisfatório para obtenção do seu perfil *cps*-RFLP. Novas cepas destes sorotipos serão testadas a fim de se obterem seus perfis experimentais.

A partir da matriz de distância par-a-par entre todos os perfis experimentais foram calculados a especificidade do banco experimental e o número de pares de sorotipos indistinguíveis entre os perfis experimentais. O dendograma das distâncias entre os perfis experimentais demonstrou que dois pares de sorotipos eram indistinguíveis (18B/18C, 11A/11F) (Figura 22). Portanto, a especificidade do método *cps*-RFLP utilizando apenas o banco de perfis experimentais foi de 87,5% para sorotipagem e de 100% para sorogrupagem. Os sorotipos 18B/18C também foram indistinguíveis pelo método PCR-RFLP descrito por Batt e colaboradores (Bentley *et al.*, 2006), e em nossa análise *in silico*. Esses sorotipos possuem o mesmo perfil de restrição *cps*-RFLP após digestão pela XhoII pois suas sequências da região *cps* diferem em apenas dois nucleotídeos (Bentley *et al.*, 2006). O par 11A/11F apresentou o mesmo perfil *cps*-RFLP e também o mesmo tamanho do locus *cps*. Não se pode descartar a hipótese de erros na sorotipagem clássica destas cepas, tendo em vista que os seus perfis *cps*-RFLP são idênticos. Estas cepas serão reencaminhadas ao Laboratório de Referência Nacional para pneumococos para serem novamente sorotipadas pela reação de Quellung. Outra consideração importante é que os cinco sorotipos do sorogrupo 11 têm seus locus *cps* geneticamente muito semelhantes, sendo agrupados em dois subgrupos. O primeiro

composto pelos sorotipos 11A, 11D e 11F. O segundo é composto pelos sorotipos 11B e 11C. Eles diferem apenas em seus genes de acetiltransferase, e provavelmente divergiram, recentemente, de um locus *cps* ancestral comum. Todos os cinco sorotipos possuem o gene *wcwT* acetiltransferase, mas os tipos 11F, 11A, 11D têm dois genes extras de acetiltransferases (*wcjE* e *wcwc*), enquanto os tipos 11B e 11C têm apenas um gene adicional de acetiltransferase (*wcwR*) (Mavroidi *et al.*, 2007). Batt e colaboradores (2005) não testaram os pares 11A/11F em seu conjunto de 46 diferentes sorotipos, indistinguíveis em nossa análise experimental, mas relataram que para os sorogrupos 12, 18, e 25, os subtipos 12F e B, 18B e C, e 25F e A não podem ser distinguidos pelo seu método PCR-RFLP (especificidade = 86,9% para sorotipagem e 100% para sorogrupagem; e sensibilidade = 100%). O par dos sorotipos 9N/9L indistinguíveis em nosso método *cps*-RFLP pôde ser distinguido na nossa simulação *in silico* pelo uso da enzima HinfI, anteriormente publicada (Bentley *et al.*, 2006).

As diferenças observadas entre os perfis obtidos experimentalmente e os calculados podem ser em parte devidas ao fato dos iniciadores utilizados na etapa de amplificação de DNA se anelarem em regiões dos genes *dexB* e *aliA* não contidas nos segmentos do locus *cps* publicados por Bentley e colaboradores (2006). Portanto, os fragmentos mais a 3' e mais a 5' observados experimentalmente devem ser maiores do que os fragmentos calculados pela digestão *in silico* do cluster *cps* sem os extremos dos genes flanqueadores *dexB* e *aliA*. Os iniciadores publicados por Bentley e colaboradores (2006) foram testados inicialmente durante a fase inicial do projeto e não resultaram na amplificação do produto do locus *cps*, como também relatado por Kilian e colaboradores (2008). Por este motivo outro par de iniciadores foi selecionado (Batt *et al.*, 2005). Também devem ser considerados os possíveis erros na estimativa dos pesos moleculares dos fragmentos a partir de suas distâncias de migração à eletroforese em géis de agarose. Além disso, fragmentos de restrição de pesos moleculares muito próximos podem co-migrar no gel de agarose. Finalmente, deve-se considerar a possibilidade de que clones de um mesmo sorotipo apresentem perfis variantes, possivelmente devido a mutações silenciosas na região *cps* que não alteraram a estrutura antigênica da CPS. Resultados semelhantes foram relatados anteriormente para sorotipos de *E. coli* e *Shigella* (Coimbra *et al.*, 2000). Pelos motivos expostos acima, o banco de dados de perfis *cps*-RFLP foi complementado com os padrões experimentais que não corresponderam aos calculados *in silico*. Nosso banco de dados de perfis *cps*-RFLP será continuamente expandido para acomodar os perfis variantes observados experimentalmente, o que contribuirá para a melhoria contínua do desempenho do método.

Os resultados da análise final do método *cps*-RFLP, cujos perfis de restrição obtidos *in silico* e experimentais foram comparados entre si com o algoritmo de alinhamento do software

MST, demonstraram que a ferramenta possui sensibilidade igual a 94,1% e especificidade de 80,2% para sorotipagem (nove pares de sorotipos indistinguíveis) e de 100% para sorogrupagem. Após retirar os perfis experimentais dos sorotipos 7F, 11A e 11D do banco de dados final, pois a sorotipagem clássica destes sorotipos precisa ser confirmada, a especificidade do método aumentou para 84,6% (sete pares de sorotipos indistinguíveis). É importante ressaltar que todas as reações cruzadas observadas para o *cps*-RFLP foram entre sorotipos do mesmo sorogrupo. Esses achados podem ser justificados, pelo fato destes sorotipos apresentarem alta similaridade entre suas sequências da região *cps*. Em resumo, neste trabalho, conseguimos sorotipar corretamente 27 dos 32 sorotipos experimentais (84,4%) testados contra o banco de dados *in silico* complementado com os perfis experimentais, assim como 100% dos sorogrupos. Os resultados de sensibilidade e especificidade para os métodos de sorotipagem descritos estão compilados na Tabela 5.

Tabela 5: Resultados da comparação entre os métodos *cps*-RFLP, PCR-RFLP e sorotipagem clássica

	Sensibilidade (%)	Especificidade	
		Sorotipagem (%)	Sorogrupagem (%)
Sorotipagem clássica ^a (n = 38 sorotipos)	-	83,5	93,4
PCR-RFLP ^b (n = 46 sorotipos)	100	86,9	100
PCR-RFLP <i>HinfI in silico</i> (n = 90 sorotipos)	-	42,2	57,7
<i>cps</i>-RFLP <i>in silico</i> (n = 90 sorotipos)	-	84,4	100
<i>cps</i>-RFLP experimental (n = 32 sorotipos)	94,1	87,5	100
<i>cps</i>-RFLP completo (n = 91 sorotipos; 90 perfis <i>in silico</i> e 32 experimentais)	-	80,2%	100
<i>cps</i>-RFLP completo, exceto os perfis experimentais 7F, 11A e 11D (n = 91 sorotipos) ^c	94,1%	84,6	100

^a Cálculos realizados com os dados publicados por Konradsen e Europe (Konradsen e Europe, 2005).

^b Cálculos realizados com os dados disponíveis por Batt e colaboradores (Coimbra *et al.*, 2000).

^c Resultado final de sensibilidade e especificidade do nosso método *cps*-RFLP.

Statens Serum Institut informou em documento referente ao painel de antissoros para sorotipagem clássica que uma de suas limitações é a possibilidade de ocorrência de reações cruzadas, sendo que para quatro dos antissoros, por vezes, não será possível chegar a sorogrupos ou sorotipos específicos. Estas reações são entre o antissoro grupo 25 com o tipo 38; antissoro tipo 29 com o tipo 35B; antissoro grupo 35 com os tipos 42 e 47F; antissoro tipo

42 com os tipos 20, 31, 33A, 35A, 35B, 35C (SSI, 2012). O método *cps*-RFLP não encontrou reação cruzada para nenhum destes pares de sorotipos mostrados acima. O fabricante dos antissoros utilizados na reação de *Quellung* relatou a possibilidade de no mínimo 10 reações cruzadas, mas a literatura relata outras reações, a maioria entre sorotipos do mesmo grupo 6A/6B (Väkeväinen *et al.*, 2001), 6A/6C (Lambertsen e Kern, 2010), mas também entre sorogrupos distintos. Um estudo de validação da sorotipagem de pneumococos realizado na Europa, em 2005, pelo *Pneumococcus Reference Laboratories*, envolvendo onze laboratórios de saúde pública também relata erros na sorotipagem (Konradsen e Europe, 2005). Os laboratórios receberam 70 cepas representando 38 sorotipos (3, 9V, 14, 18C, 20, 23F, 1, 2, 4, 5, 6B, 7F, 8, 9N, 10A, 12F, 15B, 17F, 19A, 19F, 22F, 33F, 6A, 7C, 15A, 15C, 16F, 21, 24F, 36, 38, 39, 40, 41A, 41F, 42, 48). Um total de 735 sorotipagens foram realizadas, sendo que 39 (5%) foram erradas, sete desses casos foram em nível de espécie. Dentre esses erros, 11 casos foram em nível de sorogrupos ou sorotipos, em 17 casos os sorogrupos estavam corretos, mas os sorotipos estavam errados e em quatro casos as cepas foram não-tipáveis (NT). A maioria dos erros de sorotipagem incluiu o sorotipo errado dentro do grupo correto: 7C (sorotipado como 7A, 1, 7F); 9V (9A, 9N); 9N (9L); 10A (10B, 33); 18C (18B, 18F); 19F (19A); 42 (35C, NT); e 48 (45, NT). Esses relatos demonstram que os pares indistinguíveis relatados para o método *cps*-RFLP (9N/9L, 18B/18C) também apresentam reações cruzadas à sorotipagem clássica para pneumococos e que o método *cps*-RFLP, tendo como referência o conjunto dos perfis calculados *in silico* e experimentais, apresenta maior especificidade do que a sorotipagem clássica (84,6% vs. 83,5%).

Assim como nos métodos baseados em Multiplex PCR, o método *cps*-RFLP proposto tem alto potencial para superar as dificuldades relatadas nos testes sorológicos, por também ser de fácil execução, necessitar apenas de profissional com conhecimentos básicos em biologia molecular, ser necessária apenas uma única reação de amplificação, além de possibilitar a identificação de um número maior de sorotipos em relação aos métodos baseados em Multiplex PCR. A ferramenta de sorotipagem molecular proposta aqui como alternativa a sorotipagem clássica em laboratórios que não possuem o painel de antissoros mostrou-se robusta, reprodutível e de simples execução. Acreditamos que ela também seja útil para utilização em Laboratórios de Referência equipados com instalações para execução de métodos de biologia molecular básica, além da possibilidade real da análise dos fragmentos ser automatizada no futuro, apesar de que, com o kit que utilizamos do BioAnalyzer ter apresentado resultados sub-ótimos.

A sorotipagem molecular poderá também ser uma ferramenta importante nos estudos de epidemiologia molecular por propiciar a possibilidade de detecção de perfis variantes de

um mesmo sorotipo, assim como a identificação de novos sorotipos (Coimbra, R. *et al.*, 2001). O método de extração é crítico, pois somente DNA de alta qualidade forneceu produtos viáveis para a etapa de restrição, assim como a escolha da Taq DNA polimerase. Outras enzimas para amplificação de fragmentos longos também foram testadas e não demonstraram eficácia na amplificação da região *cps* (Batt *et al.*, 2005). A técnica descrita aqui permite que um grande número de isolados sejam sorotipados no mesmo dia e em diferentes laboratórios.

O conjunto de ferramentas apresentadas neste estudo foi capaz de obter 100% dos perfis experimentais dos sorotipos presentes na vacina PCV10, indicando seu potencial para o monitoramento da eficácia da vacina 10-valente. É provável que a incidência de sorotipos não vacinais aumente, assim como ocorreu em outros países que introduziram as vacinas conjugadas em seu programa de vacinação (Rudolph *et al.*, 2013), mas é difícil prever em que grau esta substituição irá ocorrer e quais sorotipos irão prevalecer. Portanto, a ferramenta proposta neste estudo poderá ampliar a vigilância contínua das populações pneumocócicas patogênicas de maior importância.

6 CONCLUSÕES

O método de sorotipagem molecular utilizando-se a endonuclease de restrição *XhoII* foi padronizado usando uma coleção de *S. pneumoniae* previamente sorotipadas representando os 32 sorotipos mais prevalentes no estado de Minas Gerais, Brasil. O método mostrou-se robusto e reprodutível.

A sensibilidade do método foi de 94,1%, especificidade de 84,6% para sorotipagem e 100% para sorogrupagem.

A ferramenta de sorotipagem molecular para *S. pneumoniae* apresentada foi capaz de obter 100% dos perfis experimentais dos sorotipos presentes na vacina PCV10 e de distingui-los, fatos que tornam o método *cps*-RFLP uma ferramenta de monitoramento adequada para avaliar a eficácia do programa brasileiro de imunização em massa contra pneumococos.

O banco de dados de referência *cps*-RFLP construído com o desenvolvimento deste trabalho está gratuitamente disponível a toda comunidade científica através do software *web-based* MST em <http://www.cebio.org/mst>, permitindo assim que outros laboratórios de saúde pública tenham acesso a esta tecnologia.

A transferência de tecnologia do Laboratório de Referência Regional para meningites bacterianas da Funed para outros laboratórios poderá permitir a expansão do sistema de sorotipagem molecular para os outros Laboratórios de Referência da federação permitindo a descoberta de novos perfis variantes de um mesmo sorotipo que por ventura não tenham correspondentes em nosso banco híbrido de perfis *in silico* e experimentais, possibilitando assim a identificação de um número cada vez maior de sorotipos pela técnica de sorotipagem molecular desenvolvida nesta pesquisa.

O banco de dados final com os padrões *cps*-RFLP gratuitamente disponível permitirá a vigilância epidemiológica em tempo real dos sorotipos de pneumococos prevalentes no Brasil para orientar as adaptações que a vacina conjugada pneumocócica deverá sofrer para manter-se eficiente ao longo do tempo.

ANEXO I – Relação das cepas utilizadas na etapa de validação experimental.

Registro Funed	Município	UF	Origem	I D	Sorot ipo	O X A	C O	E R I	S X T	V A N	C L I	O F	R I F	T E T	MICP EN	MICC EF
883/07	Belo Horizonte	M G	LCR	S p n	11A	S	S	S	R	S	S	S	S	S	nr	nr
1070/07	Espera Feliz	M G	LCR	S p n	11F	S	S	S	R	S	S	S	S	S	nr	nr
176/08	Sete Lagoas	M G	LCR	S p n	7C	S	S	S	R	S	S	S	S	S	nr	nr
295/08	Belo Horizonte	M G	LP	S p n	1	S	S	S	R	S	S	S	S	S	nr	nr
305/08	Montes Claros	M G	SG	S p n	6A/C	S	S	S	R	S	S	S	S	S	nr	nr
317/08	Sete Lagoas	M G	LCR	S p n	34	S	S	S	S	S	S	S	S	S	nr	nr
262/08	Conceição do rio Verde	M G	LCR	S p n	6A/C	S	S	S	R	S	S	S	S	S	nr	nr
619/08	Juiz de Fora	M G	LCR	S p n	6A/C	S	S	S	R	S	S	S	S	S	nr	nr
159/09	Belo Horizonte	M G	LCR	S p n	23B	R	S	S	R	S	S	S	S	S	0.250	0.060
144/09	Uberaba	M G	LCR	S p n	35B	S	S	S	S	S	S	S	S	S	nr	nr
435/09	Belo Horizonte	M G	LCR	S p n	28A	S	S	S	S	S	S	S	S	R	nr	nr
620/09	Contagem	M G	LCR	S p n	9N	S	S	S	I	S	S	S	S	S	nr	nr
731/09	Santa Barbara	M G	LCR	S p n	7F	S	S	S	S	S	S	S	S	S	nr	nr
779/09	Curvelo	M G	LCR	S p n	15C	S	S	S	S	S	S	S	S	S	nr	nr
127/10	Unaí	M G	LCR	S p n	6C	R	S	S	R	S	S	S	S	S	0.125	0.19
148/10	Caratinga	M	LCR	S	13	S	S	S	S	S	S	S	S	S	0.016	0.016

		G		p n														
421/10	Belo Horizonte	M G	LCR	S p n	10A	S	S	S	R	S	S	S	S	S	S	S	0.023	0.016
443/10	Itaobim	M G	LCR	S p n	22F	S	S	S	S	S	S	S	S	S	S	S	0.008	0.008
585/10	Monte Santo de Minas	M G	LCR	S p n	9V	S	S	R	R	S	R	S	S	S	S	S	0.012	0.012
586/10	Nanuque	M G	LCR	S p n	18A	S	S	S	S	S	S	S	S	S	S	S	0.006	0.012
648/10	Ipatinga	M G	LCR	S p n	7C	S	S	S	R	S	S	S	S	S	S	S	0.016	0.016
029/11	Uberlândia	M G	LCR	S p n	23F	R	R	S	R	S	S	S	S	S	R		0.75	0.5
080/11	Araguari	M G	LCR	S p n	24F	S	S	S	S	S	S	S	S	S	S	S	nr	nr
84/11	Manhuaçu	M G	SG	S p n	12F	S	S	S	S	S	S	S	S	S	S	S	nr	nr
103/11	Contagem	M G	LCR	S p n	6A	S	S	S	S	S	S	S	S	S	S	S	0.008	0.008
120/11	Uberlândia	M G	LCR	S p n	18C	S	S	S	I	S	S	S	S	S	S	S	0.008	0.012
124/11	Passos	M G	LCR	S p n	19A	S	S	S	R	S	S	S	S	S	S	S	0.125	0.06
143/11	Manhuaçu	M G	LCR	S p n	6B	S	S	S	S	S	S	S	S	S	S	S	0.06	0.125
149/11	Barbacena	M G	LCR	S p n	9N	S	S	S	S	S	S	S	S	S	S	S	0,032	0.064
156/11	Uberlândia	M G	LCR	S p n	16F	S	S	S	R	S	S	S	S	S	S	S	0.06	0.06
159/11	João Monlevade	M G	LCR	S p n	3	S	S	S	S	S	S	S	S	S	S	S	0.012	0.012
79/11	Uberlândia	M G	SG	S p n	29	S	S	S	R	S	S	S	S	S	S	S	0.012	0.008
192/11	Alfenas	M	LCR	S	35B	R	R	S	R	S	R	S	S	S	R		0.75	0.38

		G		p n														
240/11	Ipatinga	M G	LCR	S p n	18B	S	S	S	R	S	S	S	S	S	S	S	0.016	0.008
254/11	Sete Lagoas	M G	LCR	S p n	19F	R	S	S	R	S	S	S	S	S	S	S	0.5	0.75
345/11	Belo Horizonte	M G	LCR	S p n	4	S	S	S	S	S	S	S	S	S	S	S	0.032	0.064
381/11	Belo Horizonte	M G	LCR	S p n	3	S	S	S	S	S	S	S	S	S	S	S	0.023	0.008
387/11	Belo Horizonte	M G	LCR	S p n	6C	S	S	S	S	S	S	S	S	S	S	S	0.012	0.012
169/11	Varginha	M G	SG	S p n	35F	S	S	S	S	S	S	S	S	S	S	S	0.012	0.016
197/11	Belo Horizonte	M G	SG	S p n	3	R	S	S	R	S	S	S	S	S	S	S	0.094	0.064
198/11	Belo Horizonte	M G	SG	S p n	14	R	S	S	R	S	S	S	S	S	S	S	2	2
490/11	Ipatinga	M G	LCR	S p n	10A	S	S	S	S	S	S	S	S	S	S	S	nr	nr
511/11	Mariana	M G	LCR	S p n	9V	S	S	S	S	S	S	S	S	S	S	S	0.064	0.125
595/11	Viçosa	M G	LCR	S p n	17F	S	S	S	S	S	S	S	S	S	S	S	0.012	0.016
649/11	Belo Horizonte	M G	LP	S p n	19A	R	S	R	R	S	R	S	S	R			4	2
ATCC49 619	-	-	-	S p n	19F	-	-	-	-	-	-	-	-	-	-	-	-	-
24/12	Divinópolis	M G	LCR	S p n	19F	R	S	R	R	S	R	S	S	S			2	1
380/12	Barbacena	M G	LCR	S p n	14	R	S	R	R	S	S	S	S	S			1	0.5
509/12	Montes Claros	M G	LCR	S p n	19F	R	S	S	S	S	S	S	S	S			4	2

Legenda: **UF:** Unidade da Federação; **MG:** Minas Gerais; **LCR:** Líquido cefalorraquidiano; **SG:** Sangue; **LP:** Líquido Pleural; **ID:** Identificação; **Spn:** *Streptococcus pneumoniae*; **R:** Resistente; **S:** Sensível; **I:** Intermediário; **OXA:** Oxacilina; **CO:** Cloranfenicol; **ERI:** Eritromicina; **SXT:** Sulfametoxazol-Trimetoprima; **VAN:** Vancomicina; **CLI:** Clindamicina; **OF:** Ofloxacina; **RIF:** Rifampicina; **TET:** Tetraciclina; **MIC PEN:** Concentração Inibitória Mínima para Penicilina; **MIC CEF:** Concentração Inibitória Mínima para Ceftriaxona; **nr:** Não realizado;

ANEXO II – Banco de dados gerado na etapa de análise *in silico* pela enzima XhoII.

S.pneumoniae_1,3773,2607,2362,2273,1912,1788,1658,1452,1088,1047,1023,441,374
S.pneumoniae_2,2996,1938,1644,1452,1088,592,483
S.pneumoniae_3,4055,2637,1911,1398,336
S.pneumoniae_4,2910,2892,1928,1429,842,761,483
S.pneumoniae_5,2016,1406,605,581,528,254
S.pneumoniae_5,2943,2402,2016,2005,1772,1311,991,898,890,711,679,654,623,605,596,552,321
S.pneumoniae_6A,3921,2643,2458,1996,1452,1088,957,920,913,424,321,294
S.pneumoniae_6B,3921,2682,2541,1452,1277,1088,931,920,913,742,321,294
S.pneumoniae_7A,2886,2553,1912,1694,1452,1300,1088,741,648
S.pneumoniae_7B,2553,1912,1585,1452,1133,1088,727
S.pneumoniae_7C,1993,1778,1687,1501,1452,1429,1133,1088,483
S.pneumoniae_7F,2915,2553,1912,1694,1452,1088,1010,741,648,394
S.pneumoniae_8,2902,1364,1139,1135,1067,592,483
S.pneumoniae_9A,3184,2870,1894,1559,909,483
S.pneumoniae_9L,2676,2610,2067,1912,1750,1686,1503,1119,965,909,421
S.pneumoniae_9N,2676,2610,2068,1912,1750,1686,1503,1119,965,909,421
S.pneumoniae_9V,3184,1894,1687,1560,1500,909,483
S.pneumoniae_10A,3293,3025,2835,1556,1321,1055,802,784,782,557,529,483,268
S.pneumoniae_10B,4274,2538,1804,1331,1154,784,309
S.pneumoniae_10C,2606,2589,2546,2479,1674,1619,1493,1055,849,512,426,278
S.pneumoniae_10F,3294,2688,2589,2565,2554,2546,849,512,426,278
S.pneumoniae_11A,4130,2389,2106,1501
S.pneumoniae_11B,4031,4009,2920,2106,1560,1136,1089
S.pneumoniae_11C,4031,3935,2127,2106,1715,1542,1136,1089,620
S.pneumoniae_11D,3336,2106,1715,1501,733
S.pneumoniae_11F,3390,2573,1688,1503,592,483,421
S.pneumoniae_12A,3478,2117,1924,1735,1732,1217,1178,1060,952,761,598,483
S.pneumoniae_12B,4206,2895,2795,2537,2321,2117,2013,1217,1178,761,592,483,286
S.pneumoniae_12F,2895,2795,2538,2321,2117,2013,1217,761,592,483,286
S.pneumoniae_13,2870,1912,1731,1278
S.pneumoniae_14,2534,2012,1912,1731,1063
S.pneumoniae_15A,2750,2536,483,415
S.pneumoniae_15B,3160,3020,2538,2511,2012,483
S.pneumoniae_15C,3160,3022,2538,2511,2012,483
S.pneumoniae_15F,2887,2776,2750,2477,1088,483,421,415
S.pneumoniae_16A,3006,2531,2029,1912,1452,1279,1088,727,511
S.pneumoniae_16F,3945,3549,2994,2892,2170,1993,1452,1088,664,336
S.pneumoniae_17A,2969,2029,1452,1331,1273,1088,910,742,724,678,658,592,483
S.pneumoniae_17F,3734,3451,2875,2873,1692,1452,1286,1273,1158,1088,727,592,483
S.pneumoniae_18A,3945,3592,2709,2682,2307,2096,1452,1088,718,401,324
S.pneumoniae_18B,3667,2537,2411,2219,2018,1912,1811,1452,1292,798,660,336,290
S.pneumoniae_18C,3667,2537,2411,2219,2018,1912,1811,1452,1292,798,660,336,290
S.pneumoniae_18F,3945,2894,2411,2181,2170,2030,1452,1292,1273,1088,727,660,336

S.pneumoniae_19A,2305,2106,1685,1555,1501,1088,925,736,716
S.pneumoniae_19B,2535,1993,1452,1088,1014,898
S.pneumoniae_19C,1985,1687,1452,1429,1378,1088,828,761,727,705,483
S.pneumoniae_19F,3307,1993,1452,1088
S.pneumoniae_20,2870,1912,1898,1640,1141
S.pneumoniae_21,3991,3257,2890,1850,1518,1380,1349,1177,798,741,483,478,290,286
S.pneumoniae_22A,4046,3479,2870,1962,1576,1452,1273,1133,1088,888,727,723,592,483,299
S.pneumoniae_22F,4046,3479,2870,1962,1576,1452,1378,1133,1088,888,727,723,592,483,299
S.pneumoniae_23A,3891,2538,2033,1993,1452,1429,1088,522,483
S.pneumoniae_23B,4257,3106,2348,2308,1502,1452,1088
S.pneumoniae_23F,2531,2284,1794,1716,1577,1500,1452,1088,858,815,483
S.pneumoniae_24A,3945,2098,1452,1273,1169,1158,1100,1088,792,742
S.pneumoniae_24B,2918,2101,2033,1452,1429,1286,1157,1088,792,483
S.pneumoniae_24F,2033,1687,1500,1452,1429,1286,1277,1158,1088,792,742,483
S.pneumoniae_25A,1826,1494,899,858,748
S.pneumoniae_25F,2003,1797,1494,899,858,700
S.pneumoniae_27,3989,2980,2769,2625,2439,2098,1452,1088,483
S.pneumoniae_28A,3182,2870,2506,2438,2437,2060,1993,1452,1088,815,664,592,483
S.pneumoniae_28F,3182,2562,2506,2438,2437,2235,2060,1467,815,664,592,483
S.pneumoniae_29,2537,2360,2354,793,289
S.pneumoniae_31,2922,2099,2010,1452,1331,1088,1010,733,724,483,467
S.pneumoniae_32A,3397,3283,3192,2964,2468,2158,1452,1347,1094,1088,762,762,601,592
S.pneumoniae_32F,3397,3283,3187,3065,2468,2158,1452,1347,1094,1088,762,762,592,496
S.pneumoniae_33A,2539,2181,1581,1388,1334,762,592,538,483,418
S.pneumoniae_33B,2892,1425,1117,1103,762,592,483,429,290
S.pneumoniae_33C,2773,2539,1838,1513,1244,592,483
S.pneumoniae_33D,2892,2846,2169,1277,1117,1103,592,483,429
S.pneumoniae_33F,2539,2181,1581,1388,1334,592,483,418
S.pneumoniae_34,2955,2789,2060,909,592,483
S.pneumoniae_35A,3187,2969,2564,2105,1651,1143,592,483
S.pneumoniae_35B,4185,3436,2969,2570,1590,833,592,483
S.pneumoniae_35C,3187,2969,2902,733,592,483
S.pneumoniae_35F,1688,1503,1429,1278,1038,940,483,454
S.pneumoniae_36,3914,3226,1912,1668,1509,1501,623,375
S.pneumoniae_37,3561,2921,2198,1935,1334,483,418
S.pneumoniae_38,1826,1731,1403,1012,858,785
S.pneumoniae_39,3676,3056,2817,2566,2385,2285,602,423
S.pneumoniae_40,1912,1716,1585,1501,1452,1117,1088,727
S.pneumoniae_41A,2918,2250,1452,1331,1088,1010,724,592,483
S.pneumoniae_41F,3561,1659,1601,1502,1452,1331,1088,1010,727,724,658,483
S.pneumoniae_42,3187,2969,2771,592,526,483
S.pneumoniae_43,2923,868,423
S.pneumoniae_44,2895,2565,2321,2115,2013,1735,1217,1060,761,598,592,483
S.pneumoniae_45,3278,2924,2765,2412,2273,1749,1452,1088,761,749,592,483

S.pneumoniae_46,3478,2570,2117,1735,1217,1178,1060,761,598,483
S.pneumoniae_47A,3557,3391,2970,2899,1517,1108,1075,733
S.pneumoniae_47F,2902,1429,1038,940,733,483,454
S.pneumoniae_48,3945,3099,3007,2736,2545,1452,1286,1158,792,742,716,372

ANEXO III – Banco de dados gerado na etapa de análise experimental pela enzima XhoII.

S.pneumoniae_E1_1,3855,2676,2445,2056,1937,1829,1688,1465,1115,1067,409,340
S.pneumoniae_E2_4,2982,2496,1971,1304,822,484
S.pneumoniae_E3_5,3938,3672,2984,2468,2054,1951,1822,1588,1019,971,901,733,703,648,624,594
S.pneumoniae_E4_6A,4046,2690,2515,1986,1488,1133,949
S.pneumoniae_E5_6B,3335,2347,2030,1758,1559,1264,1147,749
S.pneumoniae_E6_6C,3626,3007,2090,1492,1336,993,946,737,511,400
S.pneumoniae_E7_7C,3374,2046,1830,1771,1537,1477,1197,1139,538
S.pneumoniae_E8_7F,3050,2637,1959,1744,1479,1378,1132,777,688
S.pneumoniae_E9_9N,3920,3766,3476,3273,2681,2587,2099,1926,1786,1727,1517,1145,978,929,486
S.pneumoniae_E10_9V,3120,2034,1729,1594,1541,940,506,470
S.pneumoniae_E11_10A,3339,3061,2895,1569,1326,1070,811,787,590,549,502,290
S.pneumoniae_E12_11A,3989,3730,3488,3317,2124,1746,1561,522
S.pneumoniae_E13_11F,4043,3813,3616,3447,2166,1765,1573,505
S.pneumoniae_E14_12F,2916,2631,2375,2145,2038,1261,817,646,493,300
S.pneumoniae_E15_13,3610,2796,2650,1786,1060
S.pneumoniae_E16_14, 2639,2086,1951,1785
S.pneumoniae_E17_15C,3916,3447,3234,2142,1589
S.pneumoniae_E18_16F,3989,3654,3101,2933,2257,2170,2017,1538,1503,1122,703
S.pneumoniae_E19_17F,3958,3714,3033,1765,1507,1338,1219,1153
S.pneumoniae_E20_18A,4021,3752,2769,2307,2166,1507,1153,794
S.pneumoniae_E21_18B,3889,3056,2527,2317,2099,1963,1884,1493,1361,1134,1063,710
S.pneumoniae_E22_18C,3814,3020,2493,2280,2084,1940,1858,1477,1327,1122,742,689,356
S.pneumoniae_E23_19A,2611,2133,2028,1579,1108,938,721
S.pneumoniae_E24a_19F,3640,3425,3281,2014,1743,1486,1128
S.pneumoniae_E24b_19F,3673,3428,3014,1506,1333,1138,755
S.pneumoniae_E24c_19F,3716,3434,2676,2216,1510,1142
S.pneumoniae_E25_23B,4224,3137,2573,2423,1540,1029,935
S.pneumoniae_E26_23F,4273,3471,3084,2936,2601,1320,846,613,539

S.pneumoniae_E27_24F,2125,1719,1372,1300,1192,1040,798,745,489
S.pneumoniae_E28_28A,3257,3118,2482,2040,1513,1144,868,654
S.pneumoniae_E29_29,3515,3332,2993,2886,1338,863,617
S.pneumoniae_E30_34,2919,2763,2357,2056,1035,925,609,524
S.pneumoniae_E31_35B,4162,3423,2932,2602,1301,835,597,493
S.pneumoniae_E32_35F,3292,1745,1548,1444,1326,1051,965,490

Legenda: “E”: Símbolo utilizado para marcar os perfis experimentais gerados para os sorotipos específicos na análise experimental.

ANEXO IV – Tamanho em pares de bases dos produtos de amplificação da região *cps* na análise experimental.

Tamanho dos produtos em pares de bases
E1_1 = 21640 pb
E2_4 = 20884 pb
E3_5 = 26106 pb
E4_6A = 17142 pb
E5_6B = 17853 pb
E6_6C = 19610 pb
E7_7C = 23676 pb
E8_7F = 24875 pb
E9_9N = 20672 pb
E10_9V = 22860 pb
E11_10A = 20736 pb
E12_11A = 19190 pb
E13_11F = 19190 pb
E14_12F = 23625 pb
E15_13 = 19874 pb
E16_14 = 22603 pb
E17_15C = 21819 pb
E18_16F = 25993 pb
E19_17F = 22622 pb
E20_18A = 22334 pb
E21_18B = 24380 pb
E22_18C = 24367 pb
E23_19A = 20293 pb
E24a_19F = 17194 pb
E24b_19F = 21441 pb
E24c_19F = 19584 pb
E25_23B = 22334 pb
E26_23F = 18747 pb
E27_24F = 23830 pb
E28_28A = 23676 pb
E29_29 = 20013 pb
E30_34 = 18747 pb
E31_35B = 21346 pb
E32_35F = 17479 pb

Legenda: “E”: Símbolo utilizado para marcar os perfis experimentais gerados para os sorotipos específicos na análise experimental; pb: pares de bases.

ANEXO V – Resultados das análises dos perfis experimentais *cps*-RFLP testados contra o banco de dados carregado no software MST.

Registro Funed	Sorotipo	Perfis <i>cps</i>-RFLP gerados pela enzima <i>Xho</i>II	MST Sorotipo	MST Score
883/07	11A	883/07,3989,3730,3488,3317,2124,1746,1561,522	11D	15.9
1070/07	11F	1070/07,4043,3813,3616,3447,2166,1765,1573,505	11D	17.5
176/08	7C	176/08,4618,3293,2040,1832,1755,1486,1180,1144	7C	6.9
295/08	1	295/08,3855,2676,2445,2056,1937,1829,1688,1465,1115,1067,409,340	1	1.0
305/08	6A/C	305/08,4107,2673,2539,2057,1500,1143,957	6A	2.0
317/08	34	317/08,2919,2763,2357,2056,1035,925,609,524	34	6.6
262/08	6A/C	262/08,3937, 2614,2503,2008,1480, 1115,930	6A	0.6
619/08	6A/C	619/08,3920,2640,2508,2034,1467,1107,929	6A	0.5
159/09	23B	159/09,4224,3137,2573,2423,1540,1029,935	23B	4.9
144/09	35B	144/09,4236,3453,2955,2640,1315,844,601,504	35B	2.0
435/09	28A	435/09,3257,3118,2482,2040,1513,1144,868,654	28A	6.9
620/09	9N	620/09,3894,3741,3499,3251,2668,2573,2077,1918,1779,1713,1517,1145,974,929,478	9L	4.8
731/09	7F	731/09,3050,2637,1959,1744,1479,1378,1132,777,688	7A	0.0
779/09	15C	779/09,3916,3447,3234,2142,1589	22A	11.1
127/10	6C	127/10,3251,1727,1517,1461,1310,970,941,728,504	35F	3.7
148/10	13	148/10,3610,2796,2650,1786,1060	4	10.8
421/10	10A	421/10,3339,3061,2895,1569,1326,1070,811,787,590,549,502,290	10A	1.6
585/10	9V	585/10,3893,3777,3387,3266,2685,2075,1915,1777,1715,1524,1146,992,949,519	9L	7.6
586/10	18A	586/10,4021,3752,2769,2307,2166,1507,1153,794	18A	5.0
648/10	7C	648/10,3374,2046,1830,1771,1537,1477,1197,1139,538	7C	6.0
029/11	23F	29/11,4273,3471,3084,2936,2601,1320,846,613,539	35B	7.7
080/11	24F	80/11,2125,1719,1372,1300,1192,1040,798,745,489	24F	3.0
084/11	12F	84/11,2916,2631,2375,2145,2038,1261,817,646,493,300	12F	1.0

103/11	6A	103/11,4035,2678,2514,2047,1481,1124,946	6A	1.0
120/11	18C	120/11,3814,3020,2493,2280,2084,1940,1858,1477,1327,1122,742,689,356	18F	1.6
124/11	19A	124/11,3535,2752,2195,1924,1663,1161,963,747	19A	4.8
143/11	6B	143/11,3335,2347,2030,1758,1559,1264,1147,749	46	5.8
149/11	9N	149/11,3918,3755,3502,3259,2730,2570,2109,1941,1794,1741,1544,1166,999,957,490	9L	6.0
156/11	16F	156/11,3989,3654,3101,2933,2257,2170,2017,1538,1503,1122,703	16F	2.4
79/11	29	79/11,3515,3332,2993,2886,1338,863,617	42	11.9
192/11	35B	192/11,4162,3423,2932,2602,1301,835,597,493	35B	1.7
240/11	18B	240/11,3889,3056,2527,2317,2099,1963,1884,1493,1361,1134,1063,710	18F	4.8
254/11	19F	254/11,3673,3428,3014,1506,1333,1138,755	19F	2.2
345/11	4	345/11,2982,2496,1971,1304,822,484	4	7.0
387/11	6C	387/11,3788,3079,2137,1508,1364,1009,755,519,408	6C	0.5
169/11	35F	169/11,3292,1745,1548,1444,1326,1051,965,490	35F	1.4
197/11	5	197/11,3938,3672,2984,2468,2054,1951,1822,1588,1019,971,901,733,703,648,624,594	5	2.8
198/11	14	198/11,2671,2098,1965,1807	14	1.4
490/11	10A	490/11,3373,3090,2916,1576,1349,1082,824,799,649,556,506,283	10A	2.3
511/11	9V	511/11,3120,2034,1729,1594,1541,940,506,470	9V	1.2
595/11	17F	595/11,3958,3714,3033,1765,1507,1338,1219,1153	17F	5.1
649/11	19A	649/11,2611,2133,2028,1579,1108,938,721	19A	0.9
ATCC49619	19F	ATCC49619,3640,3425,3281,2014,1743,1486,1128	19F	8.4
380/12	14	380/12,2639,2086,1951,1785	14	0.8
24/12	19F	24/12,3758,3404,2993, 2087,1500,1326,1137,785	19F	5.4
509/12	19F	509/12,3716,3434, 2676,2216,1510,1142	19F	2.2

Legenda: MST: Resultado da análise no software *Molecular Serotyping Tool* (MST), MG: Minas Gerais.

ANEXO VI – Banco de dados gerado na etapa de análise experimental no sistema automatizado de eletroforese (BioAnalyzer).

Registro Funed	Sorotipo	Perfis cps-RFLP gerados pela enzima XhoII	MST Sorotipo	MST Score
149/11	9N	149/11,3465,2519,2225,1707,1557,1428,1075,932,905,440	9L	4.8
024/12	19F	24/12,3092,2558,1432,1273,1076,743	16A	3.1
120/11	18C	120/11,3135,2397,1886,1657,1490,1371,1225,1043,728,646,344	7A	2.6
143/11	6B	143/11, 3056,1940,1664,1400,1257,1194,1068,731	7A	3.4
240/11	18B	240/11,3226,2461,1936,1729,1418,1319,1062,724,643,345	7A	3.1
254/11	19F	254/11,2906,2426,1384,1203,1049,755	16A	5.7
341/11	6A	341/11,2423,1395,1197,930,752,701,490,379,293	7F	14.4
80/11	24F	80/11,2125,1719,1372,1300,1192,1040,798,745,489	24F	3.0
84/11	12F	84/11,2600,2150,1713,1296,766,595,493,300	46	6.1
149/11	9N	149/11,3420,2488,2293,1741,1435,1480,1076,932,913,436	9L	4.8
192/11	35B	192/11,4375,3708,2861,2423,1354,835,601,492	35B	6.8
197/11	5	197/11,4458,2858,2241,1944,1749,1463,964,929,708,678,641,618,602,574,555,229	5	3.1
120/11	18C	120/11,4571,3419,2594,1863,1455,1356,1096,726,647,344,210	7A	5.7
490/11	10A	490/11,4041,2878,2544,1357,1280,993,789,630,527,489,385,270	33A	10.9
511/11	9V	511/11,4160,2738,2652,725,704,694,583,520,452,304,279	12B	20.7
731/09	7F	731/09,4086,3949,2250,2091,1855,1345,996,727,686,630,556	5	13.6
80/11	24F	80/11,4105,3857,3659,2920,2450,1680,1424,1238,1850,788,501	35B	10.7
192/11	35B	192/11,4114,3332,2684,2260,1240,811,596,492	35B	7.5
120/11	18C	120/11,4415,3185,2437,1901,1723,1381,1292,1006,700,631,333	7A	1.6
490/11	10A	490/11,4184,2998,2620,1452,1337,1018,798,636,521,486,270	10A	8.8
511/11	9V	511/11,3083,1622,930,544,495,448	11F	12.8
731/09	7F	731/09,4294,2523,2417,1401,1069,751,644,594	28A	12.3
029/11	23F	29/11,4168,3258,2642,2188,1197,770,570,471	46	7.2
029/11	23F	29/11,3345,2633,1542,1239,1039,697,462,267	19F	14.7
148/10	13	148/10,4579,2598,2281,1694,976,927	19A	9.9
169/11	35F	169/11,4901,2798,1757(INCERTA),1444,1335,1215,968,915,474,437	35F	1.5
262/08	6A/C	262/08,3756,2270,1775,1465,1111,932,317,294	6A	11.6

295/08	1	295/08,4273,3843,2438,2148,1775,1481,1098,382,293	18A	9.2
345/11	4	345/11,2637,1854,1383,777,491	4	5.9
509/12	19F	509/12,2868,2174,1790,1412,1070,771,505	16A	3.9
595/11	17F	595/11,4840,3639,3118,2643,1850,1661,1446,1287,1111,1094,720,596,488	17F	3.0
649/11	19A	649/11,4425,2216,1865,1464,1130,918,748,718	19A	3.7
883/07	11A	883/07,3200,2902,2718,1964,1453,434	9A	14.5
1070/07	11F	1070/07,3595,3095,2824,2057,1513,439	9A	7.5
148/10	13	148/10,2518,2214,1511,963,919	19A	6.2
595/11	17F	595/11,3380,3000,2504,1733,1435,1225,1073,724,695,677,591,487	45	4.3
029/11	23F	29/11,3199,2562,2137,1232,807,584,482	46	5.3
148/10	13	148/10,4500,2988,2521,2179,1529,955,915	19A	2.7
159/09	23B	159/09,4133,2667,2203,1348,944,877,698,649,423	6A	14
156/11	16F	156/11,3962,2948,1884,1742,1394,1089,771,674,658,335	7A	8.2
345/11	4	345/11,3194,2976,2583,2085,1775,1370,764,485	16A	6.1
176/08	7C	176/08,4168,2774,1663,1500,1432,1143,1075,481	7C	8.8
435/09	28A	435/09,4494,2624,2185,1722,1439,1065,811,655,583,481	2.1	7.4
380/12	14	380/12,4049,2196,1755,1548,249	11C	6.8
586/10	18A	586/10,3862,3362,2404,1846,1782,1422,1060,700,395,319	18A	7.6
648/10	7C	648/10,4519,2723,1684,1444,1411,1115,473	2.1	7.5
1070/07	11F	1070/07,3641,3128,2941,2101,1532,438	27	10.8
595/11	17F	595/11,3449,3010,2549,1484,1265,1103,719,590,483	16A	5.9
144/09	35B	144/09,4378,3271,2651,2235,1279,810,588,484,426	33A	6.9
305/08	6A/C	305/08,3758,2227,1714,1453,1101,927,423,314,291	6A	10.4
317/08	34	317/08,2964,2564,2042,1755,965,902,574,474	34	3.7
585/10	9V	585/10,3241,3101,2490,2234,1802,1603,1476,1111,945,917,436	9L	5.5
620/09	9N	620/09,3612,2564,2445,1809,1440,1123,937,914,428	6A	3.1
127/10	6C	127/10,2430,1404,1217,926,749,705,488,377,314,291	6B	15.2
144/09	35B	144/09,4413,3328,2669,2242,1277,809,588,484	46	5.7
149/11	9N	149/11,3448,2504,2292,1794,1568,1460,1118,941,916,435	1	7.5
421/10	10A	421/10,3918,2795,2505,1330,1247,991,804,578,537,494,271	10A	12.8
619/08	6A/C	619/08,3874,2183,1662,1413,1106,912,426,308,285	6A	8.2
620/09	9N	620/09,3421,2441,2319,1728,1497,1418,1104,936,921,434	1	9.5

595/11	17F	595/11,3293,2458,1411,1215,1063,700,575,475	41A	7.0
345/11	4	345/11,2593,2056,1261,760,484,368	16A	13.6
435/09	28A	435/09,2658,2274,1732,1441,1091,822,654,587,493,484	2.1	5.4
586/10	18A	586/10,3896,3519,2455,1914,1771,1425,1089,699,390,314	18A	4.1

Legenda: MST: Resultado da análise no software *Molecular Serotyping Tool* (MST).

ANEXO VII – Requisito legal.



Excelência em Gestão, Patrimônio da Saúde Pública.

DECLARAÇÃO

Declaro que tenho ciência e autorizo o Dr. Dhian Renato Almeida Camargo – Analista e Pesquisador de Saúde e Tecnologia, MASP: 1098704-8, a manipular as cepas padrão de *Streptococcus pneumoniae* do Laboratório de Referência Regional para Meningites Bacterianas e colher os dados referentes a estas cepas dos livros de registro das coleções de cepas do Serviço de Doenças Bacterianas e Fúngicas da Divisão de Epidemiologia e Controle de Doenças do Instituto Octávio Magalhães, Fundação Ezequiel Dias, salvaguardando os interesses dos pacientes, sua imagem e sua privacidade. As cepas padrão e os dados coletados farão parte do projeto de pesquisa apresentado com vistas à obtenção do Título de Mestre em Ciências na área de concentração Biologia Celular e Molecular do Centro de Pesquisas René Rachou – FIOCRUZ-Minas, orientado pelo Dr. Roney Santos Coimbra Pesquisador em Saúde Pública, perfil Genômica Funcional, na FIOCRUZ-Minas (Centro de Pesquisas René Rachou).

Dr. Chequer Buffe Chamone

Coordenador da Divisão de Epidemiologia e Controle de Doenças.

Instituto Octávio Magalhães - Fundação Ezequiel Dias

Divisão de Epidemiologia e Controle de Doenças
Fundação Ezequiel Dias

Martuce A. Assunção Oliveira
Bióloga / CRB 13722-4 / MASP- 1162047-3
Chefe / Serviço de Doenças Bacterianas e Fúngicas

DECD / IOM / FUNED
17/10/2012

Rua Conde Pereira Carneiro, nº 80 - Gameleira
Belo Horizonte - MG - CEP 30.510-010
www.funed.mg.gov.br
e-mail: faleconosco@funed.mg.gov.br
Tel.: 0800 283 1980



ANEXO VIII – Resumo e o certificado de apresentação do trabalho acadêmico no XXI Congresso Latinoamericano de Microbiologia (XXI ALAM).

07/01/13 Adaltech Soluções para Evento - www.adaltech.com.br

XXI Congresso Latinoamericano de Microbiologia (XXI ALAM)

ResumoID: 1494-2

Pôster (Painel)

Evento Submissao: XXI Congresso Latino Americano de Microbiologia - ALAM

AREA: Genética de Micro-organismos e Bioinformática - Divisão D

SUB-AREA: Diversidade Microbiana

A NEW MOLECULAR SEROTYPING TOOL FOR *Streptococcus pneumoniae*

Autores Camargo, D.R.A. ^{1,3}, Pais, F.S. ², Samuel, M.L. ³, Volpini, A.C. ^{1,2}, Oliveira, G. ^{1,2}, Oliveira, M.A.A. ³, Coimbra, R.S. ^{1,2}

E-mail do primeiro autor: dhian.camargo@cpqrr.fiocruz.br

Instituição ¹ FIOCRUZ-Minas - Genomics and Computational Biology Group (Avenida Augusto de Lima, 1715, Belo Horizonte-MG, CEP: 30190-002, Brazil), ² FIOCRUZ-Minas - Center for Excellence in Bioinformatics (Rua Araguari 741, 3º andar, Belo Horizonte - MG, CEP 30190-110, Brazil), ³ FUNED-MG - Service for Bacterial and Fungal diseases (Rua Conde Pereira Carneiro, 80, Belo Horizonte - MG, CEP: 30510-010, Brazil)

Resumo:

Ninety-one *Streptococcus pneumoniae* serotypes have been described, but the conjugated anti-pneumococcal vaccine distributed by the Brazilian public health system covers only the ten most prevalent in this country. The classical pneumococcal serotyping is expensive, error prone due to cross-reactivity between similar capsular antigenic polysaccharides (CPS). We propose a new tool for molecular serotyping of *S. pneumoniae* based on restriction fragment length polymorphisms of the PCR-amplified *cps loci* (*cps*-RFLP) which encodes the enzymes responsible for CPS synthesis. We predicted the *cps*-RFLP patterns by *in silico* digesting the *cps loci* of the 91 serotypes with the endonucleases *Xho*II, or *Hin*II. The latter had already been proposed for molecular serotyping of a subset of pneumococcal serotypes. Using our previously published web service Molecular Serotyping Tool (MST – www.cebio.org/mst), restriction patterns can be aligned and their distance calculated as the sum of the penalties for the edit operations that transform one pattern into the other. MST easily distinguished each serotype-specific *cps*-RFLP pattern from all other in the database. Taking into account the median of the distances between all pairs of *cps*-RFLP patterns and the number of indistinguishable pairs for each enzyme, MST was more discriminant with *Xho*II (median= 32.8; indistinguishable pairs= 7) than *Hin*II (median=8.9; indistinguishable pairs= 24). The rare *cps*-RFLP patterns pairs that MST could not distinguish when using *Xho*II correspond to some of the serotypes with reported cross-reactivity to antisera used in classical serotyping. Molecular serotyping can be automated and does not require the use of animals, being a robust alternative to serological typing. We are validating our method comparing the predicted *cps*-RFLP patterns with those experimentally obtained for the FUNED's previously serotyped *S. pneumoniae* collection. This isolates represent the most prevalent serotypes in Minas Gerais (Brazil). Preliminary results show a strong correlation between the experimental and the predicted *cps*-RFLP patterns. The final database of *cps*-RFLP patterns will be made available freely through MST, allowing a real time epidemiological surveillance of serotype shifting, a common phenomenon after massive immunization and supporting continuous improvement of pneumococci vaccines.

Palavras-chaves: *Streptococcus pneumoniae*, Molecular serotyping, *cps*-RFLP, Web service

Agência Fomento: FIOCRUZ-Minas; FAPEMIG; CNPq



XXI ALAM

Congresso Latinoamericano
de Microbiologia

SANTOS - BRASIL

Certificado

Certificamos que o trabalho intitulado A NEW MOLECULAR SEROTYPING TOOL FOR STREPTOCOCCUS PNEUMONIAE com a autoria de: CAMARGO, D.R.A., PAIS, F.S., SAMUEL, M.L., VOLPINI, A.C., OLIVEIRA, G., OLIVEIRA, M.A.A., COIMBRA, R.S. foi apresentado na forma de pôster durante o XXI Congresso Latinoamericano de Microbiologia (XXI ALAM).

De 28 de outubro a 1 de novembro de 2012 Mendes Convention Center, Santos, SP

Prof. Dr. Carlos P. Taborda

Presidente da Associação Latinoamericana de Microbiologia

Prof. Dr. Adalberto Pessoa Junior

Presidente do XXI Congresso Latino Americano de Microbiologia



Organização e realização



ANEXO IX – Resumo e o certificado de apresentação do trabalho acadêmico no X-meeting 2013.

METHOD, SOFTWARE AND DATABASE FOR MOLECULAR SEROTYPING OF *Streptococcus pneumoniae*: A TOOL SET TO ASSESS THE EFFECTIVENESS OF VACCINATION PROGRAM IN BRAZIL

DRA Camargo^{1,4}, FS Pais³, ML Samuel⁴, AC Volpini^{2,3}, G Oliveira^{2,3}, MAA Oliveira⁴, RS Coimbra¹

¹ Biosystems Informatics, Research Center René Rachou - FIOCRUZ

² Genomics and Computational Biology Group, Research Center René Rachou - FIOCRUZ

³ Center for Excellence in Bioinformatics, Research Center René Rachou - FIOCRUZ

⁴ Service for Bacterial and Fungal diseases - FUNED-MG

Background: Ninety-one *Streptococcus pneumoniae* serotypes have been described, but the conjugated anti-pneumococcal vaccine distributed by the Brazilian public health system covers the ten most prevalent in this country. Pneumococcal serotype shifting after massive immunization is a major concern. Monitoring this phenomenon requires efficient serotyping schemes. Classical pneumococcal serotyping is expensive, centralized in a few reference centers, and error prone due to cross-reactivity between similar capsular antigenic polysaccharides (CPS). We present herein a new tool set for molecular serotyping of *S. pneumoniae* that includes: 1) a molecular biology method which produces serotype-specific *fingerprints*; 2) a database to store the reference *fingerprints*; and 3) a software to predict the serotype of clinical samples by comparing their fingerprints with those in the database. The molecular method is based on the restriction fragment length polymorphisms of the PCR-amplified *cps loci* (*cps*-RFLP) which encodes the enzymes responsible for CPS synthesis. **Results:** We predicted the *cps*-RFLP patterns of 90 serotypes with previously sequenced *cps loci* by *in silico* digesting these genomic regions with the all known endonucleases, including *XhoII*, and *HinfI*. The latter had already been proposed for molecular serotyping of a subset of pneumococcal serotypes. Using our previously published web service Molecular Serotyping Tool (MST–www.ceb.io.org/mst), the restriction patterns were aligned and their distance calculated as the sum of the penalties for the edit operations that transform one pattern into the other. Taking into account the median of the distances between all pairs of *cps*-RFLP patterns and the number of indistinguishable pairs for each enzyme, *XhoII* (mean= 36.06 ± 17.93; indistinguishable pairs= 7) was more discriminant than *HinfI* (mean=10.56 ± 6.65; indistinguishable pairs= 34). Means were significantly different ($p < 10^{-15}$) by the Wilcoxon rank-sum test. The rare *cps*-RFLP patterns pairs that MST could not distinguish when using

*Xho*II correspond to some of the serotypes with reported cross-reactivity to antisera used in classical serotyping (9L/9N, 12B/12F, 15B/15C, 18B/18C, 22A/22F, 32A/32F, 33A/33F). Published oligonucleotide sequences complementary to *dexB* and the *aliA*, conserved genes flanking the *cps loci*, were used to amplify these regions of 46 epidemiologically unrelated clinical isolates previously serotyped. They represent the 31 serotypes prevalent in Minas Gerais, Brazil. A unique amplified fragment was observed for each serotype, with sizes ranging from 17 to 26 kbp. Clearly identifiable and reproducible *cps*-RFLP patterns comprising four to 17 bands were obtained after *Xho*II digestion of the amplicons and fragments separation by electrophoresis in agarose gels. Experimental and predicted *cps*-RFLP patterns were loaded to the database linked to MST. **Conclusions:** Experimental and predicted *cps*-RFLP patterns are strongly correlated. This tool set will be made freely available allowing real time epidemiological surveillance of serotype shifting, a pre-requisite to continuous improvement of pneumococci vaccines.

Keywords: *Streptococcus pneumoniae*, Molecular serotyping, *cps*-RFLP, Web service.

Supported by: FIOCRUZ-Minas; FUNED-MG; FAPEMIG; CNPq

Acknowledgments: PDTIS-FIOCRUZ - Platform RPT04B, Bioinformatics MG



X-MEETING BSB 2013

INTERNATIONAL CONFERENCE OF THE ABC AND BRAZILIAN SYMPOSIUM ON BIOINFORMATICS

RECIFE - BRAZIL - 03-06 NOV 2013



CERTIFICATE

We certify that the poster:

METHOD, SOFTWARE AND DATABASE FOR MOLECULAR SEROTYPING OF *Streptococcus pneumoniae*: A TOOL SET TO ASSESS THE EFFECTIVENESS OF VACCINATION PROGRAM IN BRAZIL

Authors: Dhian Renato Almeida Camargo; Fabiano Sviatopolk Mirsky Pais; Barbara Rezende Pereira da Mata; Michelle Lara Samuel; Ângela Cristina Volpini; Guilherme Correa de Oliveira; Marluce Aparecida Assunção Oliveira; Roney S Coimbra

was presented at the X-meeting/BSB 2013 Conference, which was held at Mar Hotel, in Recife, PE - Brazil, from November 03rd to 6th, 2013.

Guilherme Oliveira, CEBio-Fiocruz
President ABC, X-meeting 2013

Ketia Guimarães, UFPE
SBC Delegate, BSB 2013

ANEXO X – Script em linguagem Perl utilizado para download das sequências de DNA da região cps

```
#!/usr/bin/perl
```

```
#####
```

```
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# getFastaFromNCBI.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License.
```

```
#
```

```
# getFastaFromNCBI.pl is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with restrictToMST.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####
```

```
#####
```

```
use LWP::Simple;
```

```
use Getopt::Long;
```

```
my $usage = "
```

```
$0 -i <input file> -o <output file> [ -s <separator IDs> ]
```

```
$0 -h
```

```
-i <input file>      : Input file containing identifiers
```

```
-o <output file>    : Name output file
```

```
-s <separator IDs> : Separator IDs in input file. Separators accepted: \\n, \\t, \\s and \\,
```

```
Separator default is \\n
```

```
-h                  : Help message
```

```
";
```

```
$| = 1; # forces immediate prints into files rather than the buffer.
```

```
my $inputFile;
```

```
my $outputFile;
```

```
my $separator;
```

```
my $url;
```

```
my $db;
```

```
my $search;
```

```
my $content;
```

```
my $quantLines;
```

```
my $counter;
```

```
my $percent;
```

```
my $aux = 0;
```

```
my $totalSeparator;
```

```
my @soro;
```

```

main();

sub main(){

    GetOptions ("i=s" => \$inputFile,

                "o=s" => \$outputFile,

                "s=s" => \$separator,

                "h!" => \$help);

    if ($help) {

        die $usage;

    }

    if (not(defined(\$inputFile))) {

        die "\nGive an input file name", $usage;

    }

    if (not(defined(\$outputFile))) {

        die "\nGive an output file name", $usage;

    }

    if(not(defined(\$separator))){

        $separator = "\\n";

    }

    $url = "http://www.ncbi.nlm.nih.gov/entrez/eutils";

    $db = "nucleotide";

    $quantLines = countLines(\$inputFile);

    open(IN, \$inputFile);

    open(OUT, "> \$outputFile");

    $counter = 1;

```

```

print "Progress...\n";

while(<IN>){

    if(/$separator/){

        $aux = 1;

        $sorotype = $_;

        @soro = split($separator,$sorotype);

        $totalSeparator = @soro;

        for(@soro){

            if($separator eq "\\t" or $separator eq "\\s" or $separator eq "\",
                or $separator eq ","){

                $percent = ($counter++*100)/$totalSeparator;

            }else{

                $percent = ($counter++*100)/$quantLines;

            }

            print sprintf("%.2f%%\n",$percent);

            $search = "efetch.fcgi?db=$db&id=$_&rettype=fasta";

            $content = get "$url/$search";

            print OUT $content;

        }

    }else{

        print "Separator not found in file.\n";

        while(<IN>){}

    }

}

```

```
print "Fasta file created.\n" if($aux);

close(IN);

close(OUT);

}

sub countLines{

    open(IN, $_[0]);

    $count = 0;

    while(<IN>){

        ++$count;

    }

    close(IN);

    return $count;

}
```

-Fim-

ANEXO XI – Script em linguagem Perl utilizado para identificar as endonucleases com sítios de restrição em todas as sequências de DNA da região *cps*

```
#!/usr/bin/perl
```

```
#####
```

```
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# getCommonEnzymes.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License.
```

```
#
```

```
# getCommonEnzymes.pl is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with restrictToMST.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####
```

```
#####
```

```

use LWP::Simple;

use Getopt::Long;

my $usage = "

$0 -i <input file> [ -o <output file> ]

$0 -h

-i <input file>      : Remap input file

-o <output file>    : Output name file. Output default is in terminal.

-h                  : Help message

";

$|= 1; # forces immediate prints into files rather than the buffer.

my $inputFile;

my $totalCount = 0;

my %hashTodos = ();

    GetOptions ("i=s" => \$inputFile,

                "o=s" => \$outputFile,

                "h!" => \$help);

    if ($help) {

        die $usage;

    }

    if (not(defined($inputFile))) {

        die "\nGive an input file name", $usage;

    }

open(IN, $inputFile);

while(<IN>){

```



```

if(/# Enzymes that cut Frequency Isoschizomers/){

    $totalCount++;

    while(<IN>){

        if($_ ne "\n"){

            #print $_;

            $_ =~ /(\S+)/;

            if(not exists $hashTodos{$1}){

                $hashTodos{$1} = 1;

            }else{

                $hashTodos{$1}++;

            }

        }else{

            last;

        }

    }

    #print "\n\n";

}

}

if (not(defined($outputFile))) {

    while(my ($key, $value) = each(%hashTodos)){

        if($value == $totalCount){

            print "$key -> $value \n";

        }

    }

}

```

```
}else{  
  
    open(OUT, ">$outputFile");  
  
    while(my ($key, $value) = each(%hashTodos)){  
  
        if($value == $totalCount){  
  
            print OUT "$key -> $value \n";  
  
        }  
  
    }  
  
    close(OUT);  
  
}  
  
close(IN);  
  
-Fim -
```

ANEXO XII - Script em linguagem Perl utilizado para capturar os identificadores e ordenar os fragmentos de digestão *in silico*

```
#!/usr/bin/perl
```

```
#####  
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# restrictToMST.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software  
Foundation, version 3 of the License.
```

```
#
```

```
# restrictToMST.pl is distributed in the hope that it will be useful, but WITHOUT ANY  
WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A  
PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with  
restrictToMST.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####  
#####
```

```
use LWP::Simple;
```

```

use Getopt::Long;

my $usage = "

$0 -i <input file> -f <fasta file> -t <output format type>

$0 -h

-i <input file>          : Restrict input file

-f <fasta file>         : Fasta file

-t <output format type> : Output format type. GCT and PHYLIP are accepted.

-h                      : Help message

";

$| = 1; # forces immediate prints into files rather than the buffer.

my $inputFile;

my $inputFileFa;

my $outputFile;

my $outputTableFileTrim;

my $outputTableFileComp;

my $outputType;

my $header;

my $line;

my $sequenceName;

my $serotype;

my $count;

my $i;

my @values=();

my $copy = ();

```

```

my $auxTrim = ();

my $auxCom = ();

my %hashFA = ();

my %sizeSequencesTrimmed=();

my %sizeSequencesComp=();

my %sorotypesAndSizesTrimmed=();

my %sorotypesAndSizesComp=();

GetOptions ("i=s" => \$inputFile,
            "f=s" => \$inputFileFa,
            "t=s" => \$outputType,
            "h!" => \$help);

if ($help) {
    die $usage;
}

if(not(defined($inputFile))) {
    die "\nGive an input file name",$usage;
}else{
    unless(-e $inputFile){
        die "\nInput file not found",$usage;
    }
}

if(not(defined($inputFileFa))) {
    die "\nGive an input fasta file name",$usage;
}

```

```

if(not(defined($outputType)) {
    die "\nGive an output format type",$usage;
}

$outputFile = $inputFile.".mst";

$outputType = lc($outputType);

if($outputType eq "gct"){
    $outputTableFileTrim = $inputFile.".trimmed.gct";
    $outputTableFileComp = $inputFile.".gct";
}elsif($outputType eq "phylip"){
    $outputTableFileTrim = $inputFile.".trimmed.txt";
    $outputTableFileComp = $inputFile.".txt";
}else{
    die "\nGive a valid output format type.",$usage;
}

open(IN, $inputFile);

open(OUT, ">$outputFile");

open(TABLE, ">$outputTableFileTrim");

open(TABLE_COMP, ">$outputTableFileComp");

loadingFasta();

#####

# Loading restrict file

#####

while(<IN>){
    @values = ();

```

```

if(/#\sSequence:\s(\w+)/){

    $sequenceName = $1;

    print OUT formatName($sequenceName);

}elsif(/#\s\[[a-zA-Z0-9]+\]$/{

    $auxTrim = ();

    $auxCom = ();

    $copy = ();

    while(<IN>){

        last if(/^#$/);

        $_ =~ /#\s\t(\d+)((\t\d+)+)*;/

        push(@values, $2);

        $line = $3;

        $line =~ /\t(.+)$/;

        push(@values, split(/\t/, $1));

    }

    @values = sort{$a<=$b} @values;

    $count = $#values;

    for(@values){

        #trim the bands

        if(/^\d+/{

            if($_ >= 250 and $_ <= 4300){

                push(@{ $auxTrim }, $_);

                print OUT "$_";

                if(not exists $sizeSequencesTrimmed{$_}){

```

```

        $sizeSequencesTrimmed{$_} = 1;
    }
}
else{
    push(@{$auxCom},$_);
    if(not exists $sizeSequencesComp{$_}){
        $sizeSequencesComp{$_} = 1;
    }
}
}
}
}
print OUT "\n";
$sortotypesAndSizesTrimmed{$sequenceName} = $auxTrim;
@{$scopy} = (@{$auxTrim},@{$auxCom});
$sortotypesAndSizesComp{$sequenceName} = $scopy;
for($i = 0;$i<$count;$i++){
    delete $values[$i];
}
}
}
}
printFileTabTrim();
printFileTabComp();
print "Three files were created.\n";
print " - $outputFile\n";
print " - $outputTableFileComp\n";

```



```

print " - $outputTableFileTrim\n";

close(IN);

close(OUT);

close(TABLE);

close(TABLE_COMP);

#####

# Print trimmed data in output file

#####

sub printFileTabTrim{

    my @keyssizeSequencesTrimmed = sort {$a <=> $b} keys %sizeSequencesTrimmed;

    my @sorotypes = sort {$a cmp $b} keys %sorotypesAndSizesTrimmed;

    my $ret;

    my $sorotype;

    printHeaderTrim();

    for(@sorotypes){

        $sorotype = $_;

        print TABLE formatNameMin($_);

        if($outputType eq "phylip"){

            print TABLE " ";

        }

        for(@keyssizeSequencesTrimmed){

            if($outputType eq "gct"){

                print TABLE "\t";

            }

        }

    }

}

```

```

        $ret = searchSizeTrim($sorotype,$_);

        if($ret){

            print TABLE "1";

        }else{

            print TABLE "0";

        }

    }

    print TABLE "\n";

}

}

#####

# Print trimmed header in output file

#####

sub printHeaderTrim{

    my @keyssizeSequencesTrimmed = sort {$a <=> $b} keys %sizeSequencesTrimmed;

    if($outputType eq "gct"){

        print TABLE "#1.2\n";

        print TABLE keys( %sorotypesAndSizesTrimmed )."\t".keys(
%sizeSequencesTrimmed )."\n";

        print TABLE "ID";

        for(@keyssizeSequencesTrimmed){

            print TABLE "\t$_";

        }

        print TABLE "\n";

```

```

    }else{

        print TABLE "          ".keys( %sorotypesAndSizesTrimmed )."          ".keys(
%sizeSequencesTrimmed )."\n";

    }

}

#####

# Search

#####

sub searchSizeTrim{

    $sorotype = $_[0];

    $size = $_[1];

    $auxTrim = $sorotypesAndSizesTrimmed{$sorotype};

    my $ret = 0;

    for(@{$auxTrim}){

        if($size == $_){

            $ret = 1;

            last;

        }

    }

    return $ret;

}

#####

# Print complete data in output file

#####

```

```

sub printFileTabComp{

    my %copy = (%sizeSequencesTrimmed, %sizeSequencesComp);

    my @keyssizeSequencesComp= sort {$a <=> $b} keys %copy;

    my @sorotypes = sort {$a cmp $b} keys %sorotypesAndSizesComp;

    my $ret;

    my $sorotype;

    printHeaderComp();

    for(@sorotypes){

        $sorotype = $_;

        print TABLE_COMP formatNameMin($_);

        if($outputType eq "phylip"){

            print TABLE_COMP " ";

        }

        for(@keyssizeSequencesComp){

            if($outputType eq "gct"){

                print TABLE_COMP "\t";

            }

            $ret = searchSizeComp($sorotype,$_);

            if($ret){

                print TABLE_COMP "1";

            }else{

                print TABLE_COMP "0";

            }

        }

    }

}

```

```

        print TABLE_COMP "\n";
    }
}

#####

# Print complete header in output file

#####

sub printHeaderComp{

    my %copy = (%sizeSequencesTrimmed, %sizeSequencesComp);

    my @keyssizeSequencesComp= sort {$a <=> $b} keys %copy;

    if($outputType eq "gct"){

        print TABLE_COMP "#1.2\n";

        print TABLE_COMP keys( %sorotypesAndSizesComp )."\t".keys( %copy
    )."\n";

        print TABLE_COMP "ID";

        for(@keyssizeSequencesComp){

            print TABLE_COMP "\t$_";

        }

        print TABLE_COMP "\n";

    }else{

        print TABLE_COMP " " ".keys( %sorotypesAndSizesComp )." " ".keys(
%copy )."\n";

    }

}

#####

```

```
# Verify if size exists in all serotypes
```

```
#####
```

```
sub searchSizeComp{
```

```
    $sorotype = $_[0];
```

```
    $size = $_[1];
```

```
    $auxCom = $sorotypesAndSizesComp{$sorotype};
```

```
    my $ret = 0;
```

```
    for(@{$auxCom}){
```

```
        if($size == $_){
```

```
            $ret = 1;
```

```
            last;
```

```
        }
```

```
    }
```

```
    return $ret;
```

```
}
```

```
#####
```

```
# Load file fasta in hash
```

```
#####
```

```
sub loadingFasta{
```

```
    open(FA,$inputFileFa);
```

```
    my $id = "";
```

```
    my $sero = "";
```

```
    while(<FA>){
```

```
        if(/^>gi\\d+\\w+\\(\\w+\\(d+\\)\\.\\d+\\(\\(s\\S+\\)\\)/){
```

```

        $sid = $1;

        $sero = $3;

        $sero =~ /\(serotype.+)\)/;

        $sero = $1;

        $hashFA{$sid} = $sero;

    }

}

close(FA);

}

#####

# Format name serotype in output files

#####

sub formatNameMin{

    my $sizeID;

    my $dif;

    $sid = $_[0];

    $serotype = $hashFA{$sid};

    $serotype =~ /\(S+)\s([a-zA-Z0-9]+)\)/;

    $sizeID = length($sid."_".$2);

    $dif = ($sizeID-10);

    if($dif<2){

        $sid = substr($sid, 2);

    }else{

        $sid = substr($sid, $dif);
    }
}

```

```
    }  
    return $id."_".$2;  
}  
sub formatName{  
    $id = $_[0];  
    $serotype = $hashFA{$id};  
    $serotype =~ /(\S+)\s([a-zA-Z0-9]+)/;  
    return $id."_".$2;  
}
```

-Fim -

ANEXO XII - *Script* em linguagem Perl utilizado para gerar as matrizes de distância entre os perfis de restrição

```
#!/usr/bin/perl
```

```
#####
```

```
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# runMSTinGruel.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License.
```

```
#
```

```
# runMSTinGruel.pl is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with runMSTinGruel.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####
```

```
#####
```

```
use LWP::Simple;
```

```
use Getopt::Long;
```

```
my $usage = "
```

```
$0 -i <mst input file> -o <output file> -t <output format> [-g <gruel directory>]
```

```
$0 -h
```

```
-i <mst input file>          : MST input file
```

```
-o <output file>            : Output file name
```

```
-t <output format>         : Output format type. FITCH and TAB are accepted.
```

```
-g <gruel directory>       : Directory containing gruel.
```

Default is current directory.

```
-h                          : Help message
```

```
";
```

```
$|= 1; # forces immediate prints into files rather than the buffer.
```

```
my $inputFile;
```

```
my $gruelDirectory;
```

```
my $outputFile;
```

```
my $outputType;
```

```
my %hashSero = ();
```

```
my %hashConversion = ();
```

```
my $id;
```

```
my $values = "";
```

```
my @keysComp;
```

```
my @matrix = ();
```

```
my $flag;
```

```
GetOptions ("i=s" => \$inputFile,
```

```

        "g=s" => \$gruelDirectory,
        "t=s" => \$outputType,
        "o=s" => \$outputFile,
        "h!" => \$help);

if ($help) {
    die $usage;
}

if(not(defined($inputFile))) {
    die "\nGive an input file name.", $usage;
}
else{
    unless(-e $inputFile){
        die "\nInput file not found", $usage;
    }
}

if(not(defined($outputFile))) {
    die "\nGive an output file name.", $usage;
}

if(not(defined($outputType))) {
    die "\nGive an output file name.", $usage;
}
else{
    $outputType = lc($outputType);
    if($outputType ne "fitch" and $outputType ne "tab"){
        die "\nGive a valid output type.", $usage;
    }
}

```

```

}

if(defined($gruelDirectory)) {

    if(!(-d $gruelDirectory)){

        print "Gruel directory not found.\n";

        exit;

    }else{

        unless($gruelDirectory =~ /^.+\/$){

            $gruelDirectory .= "/";

        }

    }

}

}else{

    $gruelDirectory = "";

}

#unless(-e $gruelDirectory."gruel2" and !(-d $gruelDirectory."gruel2")){

#    print "File gruel2 not found in the specified directory.\n";

#    exit;

#}

open(IN, $inputFile);

while(<IN>){

    $arrayValues = ();

    $_ =~ /([a-zA-Z0-9_]+),(.+)$/;

    $id = $1;

    $values = $2;

    $id =~ /CR(\d+)_./;

```

```

        $hashConversion{$i} = $id;

        $hashSero{$i} = $values;

    }

    close(IN);

    open(OUT, ">$outputFile");

    my @keysComp = sort {$a <=> $b} keys %hashSero;

    my $i,$j;

    my $size = @keysComp;

    print "Loading data.\n";

    print "Creating matrix";

    loadData();

    print "Print matrix.\n";

    if($outputType eq "fitch"){

        headerFitch();

        printDataFitch();

    }else{

        headerTab();

        printDataTab();

    }

    close(OUT);

    $flag = `pwd`;

    $/ = "\n";

    chomp($flag);

    print "Output file created in $flag/$outputFile\n";

```

```

#####

# functions

#####

sub printDataFitch{

    for($i=0;$i<$size;$i++){

        print OUT formatName($hashConversion{$keysComp[$i]});

        for($j=0;$j<$size;$j++){

            if($j==0){

                print OUT " ";

            }

            print OUT " $matrix[$i][$j]";

        }

        print OUT "\n";

    }

}

sub printDataTab{

    for($i=0;$i<$size;$i++){

        print OUT $hashConversion{$keysComp[$i]};

        for($n=0;$n<$i;$n++){

            print OUT "\t";

        }

        #se quiser imprimir todos no lugar do $i coloca 0

        for($j=$i;$j<$size;$j++){

            print OUT "\t$matrix[$i][$j]";

        }

    }

}

```

```

    }

    print OUT "\n";

}

}

sub loadData{

    my $count = 0;

    my $total = $size;

    my $percent;

    $path = $gruelDirectory."gruel2";

    for($i=0;$i<$size;$i++){

        $percent = ($count++*100)/$total;

        print sprintf("...%d%%", $percent);

        for($j=0;$j<$size;$j++){

            if($i<$j){

                $a = ` $path -T $hashSero{$keysComp[$i]} -E
$hashSero{$keysComp[$j]} `;

                $a =~ /#\w+;\d\.\d:(\d+\.\d+)/;

                $matrix[$i][$j] = $1;

            }elseif($i == $j){

                $matrix[$i][$j] = 0;

            }else{

                $matrix[$i][$j] = $matrix[$j][$i];

            }

        }

    }

}

```

```

    }

    print "...100%\n";
}

sub headerFitch{

    $size = @keysComp;

    print OUT "  ".$size."\n";

}

sub headerTab{

    print OUT "$size\n";

    for(@keysComp){

        print OUT "\t$hashConversion{$_}";

    }

    print OUT "\n";

}

sub formatName{

    my $sizeID;

    my $dif;

    $sid = $_[0];

    $sizeID = length($sid);

    $dif = ($sizeID-10);

    if($dif<2){

        $sid = substr($sid, 2);

    }else{

        $sid = substr($sid, $dif);

```



```
}
```

```
return $id;
```

```
}
```

-Fim-

ANEXO XIV - *Script* em linguagem Perl utilizado para gerar as figuras do tamanho dos produtos de amplificação da região *cps*

```
#!/usr/bin/perl -w
```

```
#####
```

```
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# amplificacao_simulada.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software  
Foundation, version 3 of the License.
```

```
#
```

```
# amplificacao_simulada.pl is distributed in the hope that it will be useful, but WITHOUT  
ANY WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A  
PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with  
runMSTinGruel.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####
```

```
#####
```

```
use strict;
```

```

use GD;

use Getopt::Long;

my $usage = "

$0 -i <mst database> [ -o <order file> ]

$0 -h

-i <mst input file>           : MST database input file

-o <order file>               : Order file name

                               Default is current directory.

-h                             : Help message

";

$| = 1; # forces immediate prints into files rather than the buffer.

my $help;

my $inputFile;

my $orderFile;

my $y_pixels = 350;

my $x_pixels = 2080;

my $scheme = new GD::Image($x_pixels,$y_pixels) || die;

my %hashSero = ();

my @arrayOrder = ();

my @arrayOrderBkp = ();

GetOptions ("i=s" => \$inputFile,

            "o=s" => \$orderFile,

            "h!" => \$help);

```

```

if ($help) {

    die $usage;

}

if(not(defined($inputFile))) {

    die "\nGive an input file name.", $usage;

}else{

    unless(-e $inputFile){

        die "\nInput file not found", $usage;

    }

}

#~ if(not(defined($orderFile))) {

    #~ die "\nGive an order file name.", $usage;

#}

sub func_log(){

    (my $t_y, my $size) = @_;

    return ( $t_y - (((log10($size) - log10(200)) / (log10(4370) - log10(200))) * $t_y) );

}

#read database containing fingerprints

open(DB, $inputFile);

while(<DB>){

    chomp;

    my ($ref_pattern, my @ref_bands) = split /,/;

    push(@arrayOrderBkp, $ref_pattern);

    $hashSero{ $ref_pattern } = \@ref_bands;

```

```

}

close(DB);

# Allocate colors

my $white = $scheme->colorAllocate(255,255,255);

my $black = $scheme->colorAllocate(0,0,0);

my $red = $scheme->colorAllocate(255,0,0);

my $blue = $scheme->colorAllocate(0,0,255);

# Put a black frame around de picture

#$scheme->rectangle(155,70,255,(log(($y_pixels-200))+70),$black);

$scheme->rectangle(1,1,($x_pixels-1),($y_pixels-1),$black);

#design maker bands in frame borderline

#foreach my $marker_band (200,300,400,500,564,700,1000,1500,2000,2322,4361){

foreach my $marker_band (3574,7804,9688,13282,16710,19994,22010,26718,32745){

    my $marker_y = 380 * (log (2774.1936/($marker_band + 19.161018))/log
22.199867374) + 370;

    $scheme->line(55,($y_pixels-$marker_y),60,($y_pixels-$marker_y),$black);

    $scheme->string(gdSmallFont,21,((($y_pixels-$marker_y)-7),"$marker_band",$black);

}

if(defined($orderFile)){

    open(ORDER, $orderFile);

    while(<ORDER>){

        my $line = $_;

        chomp($line);

        push(@arrayOrder, $line);
    }
}

```

```

}

close(ORDER);

my $startCol = 60;

for(@arrayOrder){

    my $arr = $hashSero{$_};

    $startCol += 20;

    $scheme->stringUp(gdSmallFont,$startCol,345,$_,$black);

    for(@{$arr}){

        my $exp_y = 380 * (log (2774.1936/($_ + 19.161018))/log
22.199867374) + 380;

        $scheme->line($startCol,($y_pixels-$exp_y),$startCol+8,($y_pixels-
$exp_y),$black);#design band from database pattern

    }

}

}else{

my $startCol = 60;

for(@arrayOrderBkp){

    my $arr = $hashSero{$_};

    $startCol += 20;

    $scheme->stringUp(gdSmallFont,$startCol,345,$_,$black);

    for(@{$arr}){

        my $exp_y = 380 * (log (2774.1936/($_ + 19.161018))/log
22.199867374) + 380;

        $scheme->line($startCol,($y_pixels-$exp_y),$startCol+8,($y_pixels-
$exp_y),$black);#design band from database pattern

```

```
        }  
    }  
}  
  
# Make sure we are writing to a binary stream  
  
binmode STDOUT;  
  
open (RES, ">$inputFile.png");  
  
print RES $scheme->png;  
  
close RES;  
  
-Fim-
```

ANEXO XV - Script em linguagem Perl utilizado para gerar as figuras com os perfis de restrição da região cps

```
#!/usr/bin/perl -w
```

```
#####
```

```
#####
```

```
# Copyright 2010 Fundação Oswaldo Cruz
```

```
#
```

```
# restricao_simulada.pl is free software: you can redistribute it and/or modify
```

```
# it under the terms of the GNU General Public License as published by the Free Software Foundation, version 3 of the License.
```

```
#
```

```
# restricao_simulada.pl is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY;
```

```
# without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.
```

```
# See the GNU General Public License for more details.
```

```
#
```

```
# You should have received a copy of the GNU General Public License along with runMSTinGruel.pl (file: COPYING).
```

```
# If not, see <http://www.gnu.org/licenses/>.
```

```
#
```

```
#####
```

```
#####
```

```
use strict;
```



```

use GD;

use Getopt::Long;

my $usage = "

$0 -i <mst database> [ -o <order file> ]

$0 -h

-i <mst input file>           : MST database input file

-o <order file>               : Order file name

                               Default is current directory.

-h                             : Help message

";

$| = 1;  # forces immediate prints into files rather than the buffer.

my $help;

my $inputFile;

my $orderFile;

my $y_pixels = 470;

my $x_pixels = 2080;

my $scheme = new GD::Image($x_pixels,$y_pixels) || die;

my %hashSero = ();

my @arrayOrder = ();

my @arrayOrderBkp = ();

GetOptions ("i=s" => \$inputFile,

            "o=s" => \$orderFile,

            "h!" => \$help);

if ($help) {

```

```

        die $usage;
    }

    if(not(defined($inputFile))) {

        die "\nGive an input file name.", $usage;

    }else{

        unless(-e $inputFile){

            die "\nInput file not found", $usage;

        }

    }

    #~ if(not(defined($orderFile))) {

        #~ die "\nGive an order file name.", $usage;

    #}

    sub func_log(){

        (my $t_y, my $size) = @_ ;

        return ( $t_y - (((log10($size) - log10(200)) / (log10(4370) - log10(200))) * $t_y) );

    }

    #read database containing fingerprints

    open(DB, $inputFile);

    while(<DB>){

        chomp;

        my ($ref_pattern, my @ref_bands) = split /,/;

        push(@arrayOrderBkp, $ref_pattern);

        $hashSero{$ref_pattern} = \@ref_bands;

    }

```

```

close(DB);

# Allocate colors

my $white = $scheme->colorAllocate(255,255,255);

my $black = $scheme->colorAllocate(0,0,0);

my $red = $scheme->colorAllocate(255,0,0);

my $blue = $scheme->colorAllocate(0,0,255);

# Put a black frame around de picture

#$scheme->rectangle(155,70,255,(log(($y_pixels-200))+70),$black);

$scheme->rectangle(1,1,($x_pixels-1),($y_pixels-1),$black);

#design maker bands in frame borderline

foreach my $marker_band (200,300,400,500,564,700,1000,1500,2000,2322,4361){

    my $marker_y = 380 * (log (2774.1936/($marker_band + 19.161018))/log
22.199867374) + 130;

    $scheme->line(50,($y_pixels-$marker_y),55,($y_pixels-$marker_y),$black);

    $scheme->string(gdSmallFont,21,((($y_pixels-$marker_y)-7),"$marker_band",$black);

}

if(defined($orderFile)){

    open(ORDER, $orderFile);

    while(<ORDER>){

        my $line = $_;

        chomp($line);

        push(@arrayOrder, $line);

    }

    close(ORDER);

```

```

my $startCol = 60;

for(@arrayOrder){

    my $arr = $hashSero{$_};

    $startCol += 20;

    $scheme->stringUp(gdSmallFont,$startCol,460,$_,$black);

    for(@{$arr}){

        my $exp_y = 380 * (log (2774.1936/($_ + 19.161018))/log
22.199867374) + 130;

        $scheme->line($startCol,($y_pixels-$exp_y),$startCol+8,($y_pixels-
$exp_y),$black);#design band from database pattern

    }

}

}else{

my $startCol = 60;

for(@arrayOrderBkp){

    my $arr = $hashSero{$_};

    $startCol += 20;

    $scheme->stringUp(gdSmallFont,$startCol,460,$_,$black);

    for(@{$arr}){

        my $exp_y = 380 * (log (2774.1936/($_ + 19.161018))/log
22.199867374) + 130;

        $scheme->line($startCol,($y_pixels-$exp_y),$startCol+8,($y_pixels-
$exp_y),$black);#design band from database pattern

    }

}
}

```

```
}  
  
# Make sure we are writing to a binary stream  
  
binmode STDOUT;  
  
open (RES, ">imagem.png");  
  
print RES $scheme->png;  
  
close RES;  
  
-Fim-
```

9 REFERÊNCIAS BIBLIOGRÁFICAS

Aanensen DM, Mavroidi A, Bentley SD, Reeves PR, Spratt BG. Predicted functions and linkage specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol.* 2007;189(21):7856-76.

Arbique JC, Poyart C, Trieu-Cuot P, Quesne G, Carvalho MaG, Steigerwalt AG, et al. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. *J Clin Microbiol.* 2004;42(10):4686-96.

Batt SL, Charalambous BM, McHugh TD, Martin S, Gillespie SH. Novel PCR-Restriction Fragment Length Polymorphism Method for Determining Serotypes or Serogroups of *Streptococcus pneumoniae* Isolates. *Journal of Clinical Microbiology.* 2005;43(6):2656-61.

Beckler E, Macleod P. THE NEUFELD METHOD OF PNEUMOCOCCUS TYPE DETERMINATION AS CARRIED OUT IN A PUBLIC HEALTH LABORATORY: A STUDY OF 760 TYPINGS. *J Clin Invest.* 1934;13(6):901-7.

Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, et al. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* 2006;2(3):e31.

Branda JA, Markham RP, Garner CD, Rychert JA, Ferraro MJ. Performance of the Vitek MS v2.0 system in distinguishing *Streptococcus pneumoniae* from nonpneumococcal species of the *Streptococcus mitis* group. *J Clin Microbiol.* 2013;51(9):3079-82.

Brandileone MC, de Andrade AL, Di Fabio JL, Guerra ML, Austrian R. Appropriateness of a pneumococcal conjugate vaccine in Brazil: potential impact of age and clinical diagnosis, with emphasis on meningitis. *J Infect Dis.* 2003;187(8):1206-12.

Brandileone MC, Vieira VS, Zanella RC, Landgraf IM, Melles CE, Taunay A, et al. Distribution of serotypes of *Streptococcus pneumoniae* isolated from invasive infections over a 16-year period in the greater São Paulo area, Brazil. *J Clin Microbiol.* 1995;33(10):2789-91.

Bratcher PE, Kim KH, Kang JH, Hong JY, Nahm MH. Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization. *Microbiology.* 2010;156(Pt 2):555-60.

Brisse S, Issenhuth-Jeanjean S, Grimont PA. Molecular serotyping of *Klebsiella* species isolates by restriction of the amplified capsular antigen gene cluster. *J Clin Microbiol.* 2004;42(8):3388-98.

Carvalho MaG, Tondella ML, McCaustland K, Weidlich L, McGee L, Mayer LW, et al. Evaluation and improvement of real-time PCR assays targeting *lytA*, *ply*, and *psaA* genes for detection of pneumococcal DNA. *J Clin Microbiol.* 2007;45(8):2460-6.

Centers for Disease Control and Prevention (CDC). Progress in introduction of pneumococcal conjugate vaccine--worldwide, 2000-2008. *MMWR Morb Mortal Wkly Rep.* 2008;57(42):1148-51.

Chatterjee DK, inventor Cloning and expressing XhoII restriction endonuclease and M. XhoII modification methylase from *Xanthomonas*. *EUA* 1994.

Coimbra RS, Grimont F, Grimont PA. Identification of *Shigella* serotypes by restriction of amplified O-antigen gene cluster. *Res Microbiol.* 1999;150(8):543-53.

Coimbra RS, Grimont F, Lenormand P, Burguière P, Beutin L, Grimont PA. Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). *Res Microbiol.* 2000;151(8):639-54.

Coimbra RS, Lenormand P, Grimont F, Bouvet P, Matsushita S, Grimont PA. Molecular and phenotypic characterization of potentially new *Shigella dysenteriae* serotype. *J Clin Microbiol.* 2001;39(2):618-21.

Coimbra RS, Artiguenave F, Jacques LS, Oliveira GC. MST (molecular serotyping tool): a program for computer-assisted molecular identification of *Escherichia coli* and *Shigella* O antigens. *J Clin Microbiol.* 2010;48(5):1921-3.

Collard JM, Alio Sanda AK, Jusot JF. Determination of pneumococcal serotypes in meningitis cases in Niger, 2003-2011. *PLoS One.* 2013;8(3):e60432.

Constenla D. The burden of pneumococcal disease and cost-effectiveness of a pneumococcal vaccine in Latin America and the Caribbean- A review of the evidence and a preliminary economic analyse. In: E. J, Valenzuela MT, editors.: Relatório disponível em www.sabin.org; 2007.

Coskun-Ari FF, Guldemir D, Durmaz R. One-step multiplex PCR assay for detecting *Streptococcus pneumoniae* serogroups/types covered by 13-valent pneumococcal conjugate vaccine (PCV13). *PLoS One.* 2012;7(12):e50406.

Dias CA, Teixeira LM, Carvalho MaG, Beall B. Sequential multiplex PCR for determining capsular serotypes of pneumococci recovered from Brazilian children. *J Med Microbiol.* 2007;56(Pt 9):1185-8.

Dubois D, Segonds C, Prere MF, Marty N, Oswald E. Identification of clinical *Streptococcus pneumoniae* isolates among other alpha and nonhemolytic streptococci by use of the Vitek MS matrix-assisted laser desorption ionization-time of flight mass spectrometry system. *J Clin Microbiol.* 2013;51(6):1861-7.

Ferreira LL, Carvalho ES, Berezin EN, Brandileone MC. [Nasopharyngeal colonization and antimicrobial resistance of *Streptococcus pneumoniae* isolated from children with acute rhinopharyngitis]. *J Pediatr (Rio J)*. 2001;77(3):227-34.

Franco-Alvarez de Luna F, Causse del Río M, Ibarra González A, Rodríguez López FC, Casal Román M. [*Streptococcus pneumoniae*: antibiotic resistance and serotypes in a two-year period]. *Rev Esp Quimioter*. 2005;18(3):217-21.

Frazão N, Brito-Avô A, Simas C, Saldanha J, Mato R, Nunes S, et al. Effect of the seven-valent conjugate pneumococcal vaccine on carriage and drug resistance of *Streptococcus pneumoniae* in healthy children attending day-care centers in Lisbon. *Pediatr Infect Dis J*. 2005;24(3):243-52.

Fundação Oswaldo Cruz (FIOCRUZ). Centro de Pesquisas René Rachou (CPQRR). PII: Programa de Incentivo à Inovação no Centro de Pesquisas René Rachou / Fundação Oswaldo Cruz. Centro de Pesquisas René Rachou. Belo Horizonte: CPQRR, 2011, p. 80.

García E, Llull D, Muñoz R, Mollerach M, López R. Current trends in capsular polysaccharide biosynthesis of *Streptococcus pneumoniae*. *Res Microbiol*. 2000;151(6):429-35.

Grimont F, Lejay-Collin M, Talukder KA, Carle I, Issenhuth S, Le Roux K, et al. Identification of a group of shigella-like isolates as *Shigella boydii* 20. *J Med Microbiol*. 2007;56(Pt 6):749-54.

Guerrero R. Bergey's manuals and the classification of prokaryotes. *Int Microbiol*. 2001;4(2):103-9.

Hathaway LJ, Brugger S, Martynova A, Aebi S, Mühlemann K. Use of the Agilent 2100 bioanalyzer for rapid and reproducible molecular typing of *Streptococcus pneumoniae*. *J Clin Microbiol*. 2007;45(3):803-9.

Henrichsen J, Robbins JB. Production of monovalent antisera by induction of immunological tolerance for capsular typing of *Streptococcus pneumoniae*. *FEMS Microbiol Lett*. 1992;73(1-2):89-93.

Holliday MG. Pneumococcal typing by polyvalent counterimmunoelectrophoresis (PIE). *J Immunol Methods*. 1981;46(2):243-9.

Henrichsen J. Six newly recognized types of *Streptococcus pneumoniae*. *J Clin Microbiol*. 1995;33(10):2759-62.

Institut SS. Pneumococcal antisera for in vitro diagnostic use 2012 [cited 2012 15 Jan. 2013]. 4th Edition:[Available from:

http://www.ssi.dk/English/SSI%20Diagnostica/Products%20from%20SSI%20Diagnostica/Antisera_antibodies/Pneumococcus%20antisera/~media/Admin/Diagnostica%20Downloads/Downloads%20UK/Packaging%20inserts/IndlgsseddelPneumococcal%20antisera%2018056%20ver%204.ashx.

Iraurgi P, Torres MJ, Gandia A, Vazquez I, Cabrera EG, Obando I, et al. Modified sequential multiplex PCR for determining capsular serotypes of invasive pneumococci recovered from Seville. *Clin Microbiol Infect.* 2010;16(9):1504-7.

Jefferies JM, Macdonald E, Faust SN, Clarke SC. 13-valent pneumococcal conjugate vaccine (PCV13). *Hum Vaccin.* 2011;7(10):1012-8.

Jiang SM, Wang L, Reeves PR. Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. *Infect Immun.* 2001;69(3):1244-55.

Jones C. Vaccines based on the cell surface carbohydrates of pathogenic bacteria. *An Acad Bras Cienc.* 2005;77(2):293-324.

Jourdain S, Drèze PA, Vandeven J, Verhaegen J, Van Melderen L, Smeesters PR. Sequential multiplex PCR assay for determining capsular serotypes of colonizing *S. pneumoniae*. *BMC Infect Dis.* 2011;11:100.

Kilian M, Poulsen K, Blomqvist T, Håvarstein LS, Bek-Thomsen M, Tettelin H, et al. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS One.* 2008;3(7):e2683.

Konradsen HB, Europe PRli. Validation of serotyping of *Streptococcus pneumoniae* in Europe. *Vaccine.* 2005;23(11):1368-73.

Kronvall G. A rapid slide-agglutination method for typing pneumococci by means of specific antibody adsorbed to protein A-containing staphylococci. *J Med Microbiol.* 1973;6(2):187-90.

Lalitha MK, Pai R, John TJ, Thomas K, Jesudason MV, Brahmadathan KN, et al. Serotyping of *Streptococcus pneumoniae* by agglutination assays: a cost-effective technique for developing countries. *Bull World Health Organ.* 1996;74(4):387-90.

Lalitha MK, Thomas K, Kumar RS, Steinhoff MC. Serotyping of *Streptococcus pneumoniae* by coagglutination with 12 pooled antisera. *J Clin Microbiol.* 1999;37(1):263-5.

Lambertsen L, Kern MB. Test of a novel *Streptococcus pneumoniae* serotype 6C type specific polyclonal antiserum (factor antiserum 6d) and characterisation of serotype 6C isolates in Denmark. *BMC Infect Dis.* 2010;10:282.

Ligozzi M, Bernini C, Bonora MG, De Fatima M, Zuliani J, Fontana R. Evaluation of the VITEK 2 system for identification and antimicrobial susceptibility testing of medically relevant gram-positive cocci. *J Clin Microbiol.* 2002;40(5):1681-6.

LUND E. Polyvalent, diagnostic pneumococcus sera. *Acta Pathol Microbiol Scand.* 1963;59:533-6.

LUND E. On the nomenclature of the pneumococcal types. *Int J Syst Bacteriol* 1970;20:321-3.

Mantese OC, Paula A, Moraes AB, Moreira TA, Guerra ML, Brandileone MC. [Prevalence of serotypes and antimicrobial resistance of invasive strains of *Streptococcus pneumoniae*]. *J Pediatr (Rio J).* 2003;79(6):537-42.

Martens P, Worm SW, Lundgren B, Konradsen HB, Benfield T. Serotype-specific mortality from invasive *Streptococcus pneumoniae* disease revisited. *BMC Infect Dis.* 2004;4:21.

Massire C, Gertz RE, Svoboda P, Levert K, Reed MS, Pohl J, et al. Concurrent serotyping and genotyping of pneumococci by use of PCR and electrospray ionization mass spectrometry. *J Clin Microbiol.* 2012;50(6):2018-25.

Mavroidi A, Aanensen DM, Godoy D, Skovsted IC, Kalsoft MS, Reeves PR, et al. Genetic relatedness of the *Streptococcus pneumoniae* capsular biosynthetic loci. *J Bacteriol.* 2007;189(21):7841-55.

Melito PL, Woodward DL, Munro J, Walsh J, Foster R, Tilley P, et al. A novel *Shigella dysenteriae* serovar isolated in Canada. *J Clin Microbiol.* 2005;43(2):740-4.

Miernyk K, Debyle C, Harker-Jones M, Hummel KB, Hennessy T, Wenger J, et al. Serotyping of *Streptococcus pneumoniae* isolates from nasopharyngeal samples: use of an algorithm combining microbiologic, serologic, and sequential multiplex PCR techniques. *J Clin Microbiol.* 2011;49(9):3209-14.

Mittman SA, Huard RC, Della-Latta P, Whittier S. Comparison of BD phoenix to vitek 2, microscan MICRoSTREP, and Etest for antimicrobial susceptibility testing of *Streptococcus pneumoniae*. *J Clin Microbiol.* 2009;47(11):3557-61.

Mond JJ, Lees A, Snapper CM. T cell-independent antigens type 2. *Annu Rev Immunol.* 1995;13:655-92.

Morais L, Carvalho MaG, Roca A, Flannery B, Mandomando I, Soriano-Gabarró M, et al. Sequential multiplex PCR for identifying pneumococcal capsular serotypes from South-Saharan African clinical isolates. *J Med Microbiol.* 2007;56(Pt 9):1181-4.

Mudany MA, Kikuchi K, Totsuka K, Uchiyama T. Evaluation of a new serotyping kit for *Streptococcus pneumoniae*. *J Med Microbiol*. 2003;52(Pt 11):975-80.

Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol*. 1970;48(3):443-53.

Neufeld F. Ueber die Agglutination der Pneumokokken und iiber die Theorieen der Agglutination. *Z Hyg Infektionskr*. 1902;40:54-72.

Novaes HM, Sartori AM, Soárez PC. Hospitalization rates for pneumococcal disease in Brazil, 2004 - 2006. *Rev Saude Publica*. 2011;45(3):539-47.

O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet*. 2009;374(9693):893-902.

Organización Panamericana de la Salud (OPS). Informe Regional de SIREVA II, 2011: datos por país y por grupos de edad sobre las características de los aislamientos de *Streptococcus pneumoniae*, *Haemophilus influenzae* y *Neisseria meningitidis* en procesos invasores. Washington, D.C.: OPS, 2012, p. 62-85. ISBN 978-92-75-31701-3.

Organización Panamericana de la Salud (OPS). Informe Regional de SIREVA II, 2012: datos por país y por grupos de edad sobre las características de los aislamientos de *Streptococcus pneumoniae*, *Haemophilus influenzae* y *Neisseria meningitidis* en procesos invasores. Washington, D.C.: OPS, 2013, p. 37-64. ISBN 978-92-75-31763-1.

Pai R, Gertz RE, Beall B. Sequential multiplex PCR approach for determining capsular serotypes of *Streptococcus pneumoniae* isolates. *J Clin Microbiol*. 2006;44(1):124-31.

Park IH, Park S, Hollingshead SK, Nahm MH. Genetic basis for the new pneumococcal serotype, 6C. *Infect Immun*. 2007;75(9):4482-9.

Park MK, Briles DE, Nahm MH. A latex bead-based flow cytometric immunoassay capable of simultaneous typing of multiple pneumococcal serotypes (Multibead assay). *Clin Diagn Lab Immunol*. 2000;7(3):486-9.

Pichon B, Ladhani SN, Slack MP, Segonds-Pichon A, Andrews NJ, Waight PA, et al. Changes in molecular epidemiology of *streptococcus pneumoniae* causing meningitis following introduction of pneumococcal conjugate vaccination in England and Wales. *J Clin Microbiol*. 2013;51(3):820-7.

Pletz MW, Maus U, Krug N, Welte T, Lode H. Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaptation of the species. *Int J Antimicrob Agents*. 2008;32(3):199-206.

Raymond F, Boucher N, Allary R, Robitaille L, Lefebvre B, Tremblay C, et al. Serotyping of *Streptococcus pneumoniae* Based on Capsular Genes Polymorphisms. *PLoS One*. 2013;8(9):e76197.

Reinert RR, Paradiso P, Fritzell B. Advances in pneumococcal vaccines: the 13-valent pneumococcal conjugate vaccine received market authorization in Europe. *Expert Rev Vaccines*. 2010;9(3):229-36.

Richter SS, Howard WJ, Weinstein MP, Bruckner DA, Hindler JF, Saubolle M, et al. Multicenter evaluation of the BD Phoenix Automated Microbiology System for antimicrobial susceptibility testing of *Streptococcus* species. *J Clin Microbiol*. 2007;45(9):2863-71.

Rudolph K, Bruce MG, Bulkow L, Zulz T, Reasonover A, Harker-Jones M, et al. Molecular epidemiology of serotype 19A *Streptococcus pneumoniae* among invasive isolates from Alaska, 1986-2010. *Int J Circumpolar Health*. 2013;72.

Sanz JC, Culebras E, Ríos E, Rodríguez-Avial I, Wilhelmi I, Ramos B, et al. Direct serogrouping of *Streptococcus pneumoniae* strains in clinical samples by use of a latex agglutination test. *J Clin Microbiol*. 2010;48(2):593-5.

Selva L, del Amo E, Brotons P, Muñoz-Almagro C. Rapid and easy identification of capsular serotypes of *Streptococcus pneumoniae* by use of fragment analysis by automated fluorescence-based capillary electrophoresis. *J Clin Microbiol*. 2012;50(11):3451-7.

Shaaly A, Tellevik MG, Langeland N, Høiby EA, Jureen R. Comparison of serotyping, pulsed field gel electrophoresis and amplified fragment length polymorphism for typing of *Streptococcus pneumoniae*. *J Med Microbiol*. 2005;54(Pt 5):467-72.

Shutt CK, Samore M, Carroll KC. Comparison of the Denka Seiken slide agglutination method to the quellung test for serogrouping of *Streptococcus pneumoniae* isolates. *J Clin Microbiol*. 2004;42(3):1274-6.

Slotved HC, Kaltoft M, Skovsted IC, Kerrn MB, Espersen F. Simple, rapid latex agglutination test for serotyping of pneumococci (Pneumotest-Latex). *J Clin Microbiol*. 2004;42(6):2518-22.

Smirnova LF, Avtushenko SS. [Mechanisms of the protective action of cryoprotectors on *Escherichia coli* cells]. *Mikrobiologiya*. 1988;57(3):494-8.

76. Sørensen UB. Typing of pneumococci by using 12 pooled antisera. *J Clin Microbiol*. 1993;31(8):2097-100.

Vickers I, O'Flanagan D, Cafferkey M, Humphreys H. Multiplex PCR to determine *Streptococcus pneumoniae* serotypes causing otitis media in the Republic of Ireland with further characterisation of antimicrobial susceptibilities and genotypes. *Eur J Clin Microbiol Infect Dis*. 2011;30(3):447-53.

Vieira AC, Gomes MC, Rolo Filho M, Eudes Filho J, Bello EJ, de Figueiredo RB. *Streptococcus pneumoniae*: a study of strains isolated from cerebrospinal fluid. *J Pediatr (Rio J)*. 2007;83(1):71-8.

Väkeväinen M, Eklund C, Eskola J, Käyhty H. Cross-reactivity of antibodies to type 6B and 6A polysaccharides of *Streptococcus pneumoniae*, evoked by pneumococcal conjugate vaccines, in infants. *J Infect Dis*. 2001;184(6):789-93.

Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, Lynfield R, et al. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. *N Engl J Med*. 2003;348(18):1737-46.

World Health Organization (WHO). Pneumococcal conjugate vaccine for childhood immunization--WHO position paper. *Wkly Epidemiol Rec*. 2007;82(12):93-104.

Yu J, Lin J, Kim KH, Benjamin WH, Nahm MH. Development of an automated and multiplexed serotyping assay for *Streptococcus pneumoniae*. *Clin Vaccine Immunol*. 2011;18(11):1900-7.

Yun KW, Cho EY, Hong KB, Choi EH, Lee HJ. *Streptococcus pneumoniae* type determination by multiplex polymerase chain reaction. *J Korean Med Sci*. 2011;26(8):971-8.