ALICE TAVARES REIS ANDRADE

“Evolução do gene *sodC* nas bactérias naturalmente transformáveis *Neisseria meningitidis* e *Haemophilus influenzae*”.

Campinas, 2013

SECRETARIA
DE
PÓS-GRADUAÇÃO
I. B.

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

ALICE TAVARES REIS ANDRADE

"EVOLUÇÃO DO GENE *sodC* NAS BACTÉRIAS NATURALMENTE TRANSFORMÁVEIS *Neisseria meningitidis* E *Haemophilus influenzae*."

Este exemplar corresponde à redação final da tese defendida pelo(a) candidato (a) Alice Tavares Reis Andrade e aprovada pela Comissão Julgadora.

Dissertação apresentada ao Instituto de Biologia para obtenção do Título de Mestra em Biologia Funcional e Molecular, na área de Bioquímica.

Orientador: Prof. Dr. Marcelo Lancellotti

CAMPINAS
2013

FICHA CATALOGRÁFICA ELABORADA POR
MARA JANAINA DE OLIVEIRA – CRB8/6972
BIBLIOTECA DO INSTITUTO DE BIOLOGIA - UNICAMP

An24e

Andrade, Alice Tavares Reis, 1977-
Evolução do gene sodC nas bactérias naturalmente transformáveis *Neisseria meningitidis* e *Haemophilus influenzae* / Alice Tavares Reis Andrade. – Campinas, SP: [s.n.], 2013.

Orientador: Marcelo Lancellotti.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Transferência genética horizontal. 2. Seleção natural. 3. *Neisseria meningitidis*. 4. *Haemophilus influenzae*. 5. Superóxido dismutase. I. Lancellotti, Marcelo, 1976-. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Informações para Biblioteca Digital

Título em Inglês: Evolution of the sodC gene in the naturally transformable bacteria *Neisseria meningitidis* and *Haemophilus Influenzae*
Palavras-chave em Inglês:
Horizontal gene transfer
Natural selection
Neisseria meningitidis
Haemophilus influenzae
Superóxido dismutase
Área de concentração: Bioquímica
Titulação: Mestra em Biologia Funcional e Molecular
Banca examinadora:
Marcelo Lancellotti [Orientador]
Sérgio Furtado dos Reis
Fabiana Fantinatti Garboggini
Data da defesa: 28-01-2013
Programa de Pós Graduação: Biologia Funcional e Molecular

Campinas, 28 de janeiro de 2013

BANCA EXAMINADORA

Prof. Dr. Marcelo Lancellotti (Orientador)

Assinatura

Prof. Dr. Sergio Furtado dos Reis

Assinatura

Dra. Fabiana Fantinatti Garboggini

Assinatura

Profa. Dra. Fernanda Ramos Gadelha

Assinatura

Dr. Marcus Alexandre Finzi Corat

Assinatura

"A ignorância gera confiança com mais freqüência do que o conhecimento."

Charles Darwin

A Gustavo.

Agradecimentos

"Na sobrevivência dos indivíduos e raças favorecidas, durante a luta constante e recorrente pela existência, vemos uma poderosa e incessante forma de seleção." (Charles Darwin, em A Origem das Espécies). Como traduzir em palavras o coração humano? O fascínio, a gana, o desejo ardente e inexplicável disso que as pessoas tentam traduzir como evolução dos seres e que é tão somente o mundo onde o meu coração habita e se sente tão em paz!

Agradeço ao CNPq e a Fapesp pelo apoio financeiro ao projeto.

Agradeço a Gustavo, meu marido, companheiro, melhor amigo, confidente, minha paz, meu equilíbrio. Quero te agradecer por você ser único, e por ter esta singularidade de embarcar nos meus sonhos e nos meus anseios como se fossem os seus próprios. Por ser a minha força, quando estou exaurida, a minha serenidade quando a vida insiste em não me dar sossego, a minha alegria quando estou triste. Quero também agradecer a minha mãe por estar ao meu lado sempre e incondicionalmente! Por ser só amor, e por me dar o aconchego e o carinho de que tanto preciso!

Também quero agradecer especialmente a Renato Orsi, a quem conheço muito bem por Skype, sem praticamente tê-lo visto pessoalmente. Agradeço a sua coragem e ousadia em me ensinar a usar o PAML à distância e virtualmente, e por tudo que me ensinou sobre evolução molecular, filogenia, pseudogenes e recombinação. Sem você, esta dissertação definitivamente não teria dado certo. Obrigada pela sua paciência, humildade, inteligência, integridade, pelo seu saber tão vasto, pelo seu carater e pela sua pontualidade britânica (!). Agradeço ao Sérgio por tudo com me ensinou, pelas nossas conversas extremamente

agradáveis e pelo nosso gosto comum por chocolates incomuns (!), ao Carlos Tomate por sempre me socorrer com este troço chamado computador e a Carol Lemes por ter pacientemente me ajudado nos meus primeiros passos no BioEdit.

Como não agradecer aos meus gatos Robinho, Pandora e Millie, meus pacientes e fiéis companheiros de jornada!

Agradeço também a meu pai, pelo exemplo de força, persistência e superação, aos meus tios Flávio, Leninha, Aida e Alice, ao meu irmão Marcos, minha cunhada Cristiane e às minhas três sobrinhas lindas: Bia, Lulu e Kaká. E por fim, mas não menos importante, à minha nova e maravilhosa família: Lena, WA, Guiga, Liz, Marai, Juli, Sérgio, Bigu, Luiza, Fred, Pedro, D. Miriam, Dudé, tia Tereza, Waldemir, Cristina, Martinha (por todas as suas orações fervorosas e tantas velas acesas!), Alfredo, Daniel, Gabriela, Ana Luiza, Maria, Marluce, Lia, Zeni, Margarida, os Margotto, os Setenta, Inalda, Otília, Carlinhos, Conceição...enfim, muito obrigada pelo carinho!

"Minha meta é navegar além do por do sol. Embora não tenhamos a força que antigamente movia céus e terra, o que nós somos, nós somos. Uma boa índole e corações heróicos, enfraquecidos pelo tempo, mas fortes na vontade de lutar, procurar, achar e NÃO DESISTIR NUNCA!" (Sociedade dos Poetas Mortos, filme)

Resumo

Em 1998, foi relatada a transferência lateral do gene *sodC* do gênero *Haemophilus* para a espécie *Neisseria meningitidis*. Sabe-se que, nestes dois grupos a dinâmica deste gene é bastante distinta. Este trabalho tem por objetivo estimar árvores filogenéticas que possam apontar qual a espécie do gênero *Haemophilus* compartilhou o gene *sodC* com a espécie *N. meningitidis*. Testes de seleção positiva foram empregados no intuito de avaliar quais forças evolutivas estão sujacentes ao processo de diversificação molecular do gene nestas espécies ao longo do tempo. Além disso, foi realizada uma modelagem protéica computacional por homogia para avaliar quais substituições de aminoácidos tinham impacto no processo adaptativo da enzima nas espécies consideradas. Ao se reconstruir uma filogenia para o gene *sodC*, foi constatado que a origem deste gene na espécie *H. influenzae* é distinta. Um grupo de linhagens recebeu o gene, provavelmente por transferência lateral, da espécie *H. haemolyticus*, enquanto o outro grupo recebeu o gene da espécie *H. parainfluenzae.* Neste grupo, o gene sofreu pseudogeneização. Foi observado também que as sequências de *N. meningitidis* agrupam com as sequências que compartilham um ancestral comum com a espécie *H. haemolyticus*, porém as sequências do meningococo formam um ramo distinto dentro deste clado. Dada a alta clonalidade das sequências de *N. meningitidis,* foi constatado que o evento de transferência lateral de genes foi muito recente na escala do tempo.O teste de seleção positiva demonstrou que seleção positiva está atuando especificamente no ramo da árvore que compartilha um ancestral comum com a espécie *H. haemolyticus,* através da modificação de uma alanina por uma serina na posição 72, embora a nota geral da árvore tenha sido menor que 1. Sabe-se que pseudogenes, por não codificarem uma proteína ativa e, portanto, por não estarem sob nenhum tipo de restrição funcional, estão sob uma ação maior da deriva genética. Portanto,

diferentes forças evolutivas estão governando a evolução deste gene nas espécies consideradas. A modelagem protéica concluiu que tal modificação contribuiu para o aumento do potencial redox do sítio ativo. Desta forma, a ação da seleção positiva sob um único resíduo de aminoácido foi benéfica para a função da enzima como um todo.

Abstract

In 1998, it was reported the lateral transfer of the *sodC* gene from the genus *Haemophilus* to *Neisseria meningitidis*. It is known that this two groups show a quite distinct dynamics of this gene. This study aims to estimate phylogenetic trees that migh tpoint to which species of the genus *Haemophilus* shared the *sodC* gene with *N. meningitidis*. In addition, tests of positive selection were employed in order to assess which evolutionary forces are governing the process of molecular diversification of the gene in these species through time. Moreover, we performed a computational protein modeling by homology to asses which amino acids substitutions had an impact on the adaptative process of the enzyme in the species considered. A phylogeny of the *sodC* gene was reconstructed and it was found that this gene in *H. influenzae* has two different origins.A group of lineages has received the gene, probably by lateral transfer, from *H. haemolyticus*, whereas the other group has received the gene from *H. parainfluenzae*. In the latter, the gene has become a pseudogene.It was also observed that the sequences from *N. meningitides* group together with those sequences that share a common ancestor with *H. haemolyticus*, but they form a distinct branch within this clade. Given the high clonality of the sequences from *N. meningitidis,* it was found that the lateral gene transfer event is very recent in the time scale. A test of positive selection showed that positive selection is acting specifically in the branch that shares a common ancestor with *H. haemolyticus* through the substitution of an alanine to a serine at position 72, though the overall score of the tree is less than one. It is known that pseudogenes do not encode active proteins and therefore they are not under any kind of functional constraints, so they are under greater influence of genetic drift. Thus, it was concluded that different forces are driving the evolution of this gene in the species

considered here. Protein modeling concluded that this modification contributed to the increase in the redox potencial of the active site. Thus the action of positive selection under a single amino acid residue was beneficial to the function of the enzyme as whole.

Lista de Abreviaturas

AA: Aminoácido

ATCC: *American Type Culture Collection*

ATP: Adenina trifosfato

BHI: *Brain Heart Infusion*

CO_2: Dióxido de carbono

DNA: Ácido desoxirribonucleico

dN: Taxa de substituições Não Sinônimas

dS: Taxa de Substituições Sinônimas

dNTP: Desoxinucleosídeo Trifosfato

DO: Densidade ótica

DPOC: Doença Pulmonar Obstrutiva Crônica

ELISA: *Enzyme-Linked ImmunoSorbent Assay*

Kb: Kilobases

LCR: Líquido Céfalorraquidiano

LOS: Lipooligossacarideo

LPS: Lipopolissacarídeo

MAbs: Anticorpos Monoclonais

Mb: Megabase

NAD: Nicotinamida Adenina Dinucleotídeo

NJ: *Neighbor Joining* (Junção de Vizinhos)

NTHi: *Non Typeable Haemophilus influenzae*

OMP: *Outer Membrane Protein* (Proteína de Membrana Externa)

ORF: *Open Reading Frame* (Matriz de Leitura Aberta do Gene)

Pb: Pares de bases

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

RNA: Ácido ribonucléico

RNS: *Reactive Nitrogen Species* (Intermediários Reativos do Nitrogênio)

ROS: *Reactive Oxygen Species* (Intermediários Reativos do Oxigênio)

SOD: Superóxido Dismutase

TLG: Transferência lateral de genes

TM: Temparatura de Melting (anelamento)

US: *Uptake Sequences* (Sequências Uptake)

Lista de Tabelas

Tabela 1. Composição Química da Cápsula Polissacarídica dos Principais Sorogrupos de *N. meningitidis*.....4

Tabela 2. Classificação da espécie *H. influenzae* segundo a presença e a tipagem sorológica da cápsula polissacarídica.....9

Tabela 3.Linhagens de *N. meningtidis* utilizadas neste trabalho.....41 e 42

Tabela 4. Linhagens de *H. influenzae* utilizadas neste trabalho.....42,43 e 44

Tabela 5: Composição do tampão CTAB.....45

Tabela 6. Primers utilizados na Reação de PCR e na Reação de Sequenciamento.....46

Tabela 7: Condições padrão da reação de PCR.....47

Tabela 8: Temperatura de anelamento para cada espécie bacteriana.....47

Tabela 9: Linhagens bacterianas sequenciadas à partir do LABIOTEC.....47 e 48

Tabela 10: Números de acesso (AC) utilizados neste trabalho.....49

Tabela 11: Métodos para se estimar árvores filogenéticas e suas estratégias.....53

Tabela 12: Modelos de sítio e seus respectivos parâmetros.....58

Tabela 13: Substituições não sinônimas entre as sequências do gene *sodC* de *N. meningitidis*.....74

Tabela 14: Resultados das análises de seleção pelos modelos de sítio do programa CODEML.....102

Tabela 15: Resultados das análises de seleção pelos modelos de ramos do programa CODEML.....103

Tabela 16: Resíduos de aminoácidos com probabilidade de estarem sob seleção positiva.....103

Lista de Figuras

Figura1. Fotomicrografia eletrônica de corte de meninge humana com *N. meninigitidis*.....1

Figura 2. Fluxograma Representativo da Aquisição de *N. meningitidis*e do Desenvolvimento de Doença ou Estado de Portador.....2

Figura 3A. Distribuição Mundial dos Principais Sorogrupos de *N. meningitidis.*.....5

Figura 3B. Representação do Cinturão Africano da Meningite.....5

Figura 4. Representação da Membrana Externa do Meningococo.....7

Figura 5. *Haemophilus influenzae* corado pelo método de gram.....8

Figura 6. *H. influenzae* não capsulado corado pelo método de gram.....10

Figura 7: Representação esquemática dos mecanismos de Transferência Lateral de Genes.....13

Figura 8: Árvore filogenética evidenciando um ramo anômalo.....15

Figura 9A: Representação esquemática do sistema de secreção bacteriana do tipo II em *Klebsiella oxytoca*.....17

Figura 9B: Comparação do sistema de secreção tipoIV e do complexo protéico associado ao estado natural de competência em *Neisseria gonorrhoeae*.....17

Figura 10: Diferenças nas etapas de transformação em bactérias gram-positivas e gram-negativas.....19

Figura 11: Representação esquemática do complexo protéico responsável pela transformação em bactérias gram-negativas.....21

Figura 12A: Representação de uma paisagem com picos e vales adaptativos.....25

Figura12B: Os pontos indicam indivíduos de uma população ocupando picos e vales adaptativos.....25

Figura 13: Representação esquemática da atuação da Sod nas etapas iniciais de reações citoprotetoras.....27

Figura 14A: Estrutura cristalográfica da SodC de *N. meningitidis,* compostas por dois monômeros associados por pontes dissulfeto intramoleculares.....28

Figura 14B: Estrutura cristalográfica de um dos quatro monômeros da SodC bovina.....28

Figura 15: Representação esquemática do espaço periplasmático de bactérias gram-negativas, localização subcelular da enzima SodC.....31

Figura 16: Classificação dos Filos pertencentes ao domínio *Bacteria*.....34

Figura 17: Etapas de um alinhamento prograssivo.....52

Figura18: Possibilidades de mudança de estado do codon.....56

Figura 19: Árvore 1 para o gene *sodC* de *N. meningitidis:* somente com as 15 amostras do LABIOTEC.....66

Figura 20: Alinhamento parcial de 15 sequências de nucleotídeos correspondentes ao gene *sodC*de *N. meningitidis.*.....68

Figura 21: Árvore 2 para o gene *sodC* de *N. meningitidis.*.....71

Figura 22: Alinhamento 2, com 30 sequências do gene *sodC*de *N. meningitidis.*.....72

Figura 23: Alinhamento 3, contendo todas as sequências disponíveis para o gene *sodC* de *N. meningitidis.*.....76

Figura 24: Árvore do gene *sodC* de *N. meningitidis* com todas as sequências disponíveis no Genebank mais as sequências do LABIOTEC.....77

Figura 25: Árvore filogenética para o gene *sodC* de espécies do gênero *Haemophilus*.....81

Figura 26: Mesma árvore anterior, destancando os ramos que compartilham um ancestral comum com *H. haemolyticus* e com *H. parainfluenzae*.....84

Figura 27: Alinhamento parcial das sequências que compartilham um ancestral comum com *H. haemolyticus.*.....86

Figura 28: Alinhamento parcial das sequências que compartilham um ancestral comum com *H. parainfluenzae.*.....86

Figura 29: Alinhamento parcial das sequências AAr27, Mr27, I8809, C1421 (pseudogenes), HiB e HiF com as sequências de *H. parainfluenzae* e 2632_NTHi (genes verdadeiros).....88

Figura 30: Alinhamento parcial das sequências P2321_NTHi, O1421_NTHi, O0521_NTHi e HiE (pseudogenes) com as sequências *H. parainfluenzae* T3T1 e 26323_NTHi (genes verdadeiros).....89

Figura 31: Sequência parcial da proteína SodC de HiB. Notar o aminoácido Tirosina destacado na posição 82.....91

Figura 32: Árvore filogenética para as sequência do gênero *Haemophilus* e da espécie *N. meningitidis.*.....98

Figura 33: Alinhamento parcial de sequências representativas das espécies *N. meningitidis* (destacadas com um parênteses vermelho), *H. influenzae* e *H. haemolyticus*.....99

Figura 34: Árvore filogenética utilizada como arquivo de entrada para o programa CODEML.....102

Figura 35: Sequência primária da SodC de *N. meningitidis*.....110

Figura 36: Modelagem computacional das repercussões da mutação alanina→serina sobre a estrutura da enzima SodC.....111

Índice

1. Introdução.....1

1.1 Neisseria meningitidis.....1

1.2 *Haemophilus influenzae*.....7

1.3 A Transferência Lateral de Genes em Procariotos.....11

1.4 O Estado de Competência Natural ao Processo de Transformação Bacteriana..16

1.5 A Enzima [Cu, Zn] Superóxido Dismutase.....25

1.6 A Transferência Lateral do Gene *sodC*.....32

2. Justificativa.....38

3. Objetivos.....39

3.1 Objetivo Geral.....39

3.2 Objetivos Específicos.....39

4. Material e Métodos.....40

4.1 Material.....40

4.1.1 Meios de Cultura.....40

4.1.1.1 Infusão de cérebro-coração (BHI) e BHI Ágar Suplementada.....40

4.1.1.2 Ágar Chocolate.....40

4.1.2 Linhagens Bacterianas e Condições de Crescimento.....40

4.2 Métodos.....45

4.2.1 Extração de DNA.....45

4.2.2 Amplificação do Gene de Interesse por PCR.....45

4.2.3 Reação de Sequenciamento do Produto de PCR.....47

4.2.4 Obtenção de Sequências Homólogas à partir do GeneBank.....48

4.2.5 Alinhamento Múltiplo de Sequências de Nucleotídeos.....50

4.2.6 Estimação e Reconstrução de Árvores Filogenéticas para o gene *sodC*.....52

4.2.7 Análise da Ocorrência de Seleção Positiva (Teste de Seleção).....55

4.2.8 Modelagem Proteica Computacional por Homologia.....59

5. Resultados e Discussão.....61

5.1 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC*na espécie *N. meningitidis*.....61

5.2 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC* no gênero *Haemophilus*.....79

5.3 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC* no gênero *Haemophilus* e na espécie *N. meningitidis*.....94

5.4 Teste de Seleção Positiva.....99

5.5 Modelagem Proteica Computacional por Homologia.....108

6.Conclusões.....113

7.Referências Bibliográficas.....117

1.Introdução

1.1 *Neisseria meningitidis*

Neisseria meningitidis ou meningococo é um diplococo gram-negativo capsulado, não esporulado e não flagelado, que pertence à classe das β-Proteobactérias, à família *Neisseriaceae*, e ao gênero *Neisseria.* Juntamente com o gonococo *(Neisseria gonohoeae),* representam os únicos membros patogênicos deste gênero. *N. meningitidis* é um membro comensal facultativo das vias aéreas superiores, colonizandoestritamente o hospdeiro humano,principalmente naso e orofaringe (Tortora, 2000).

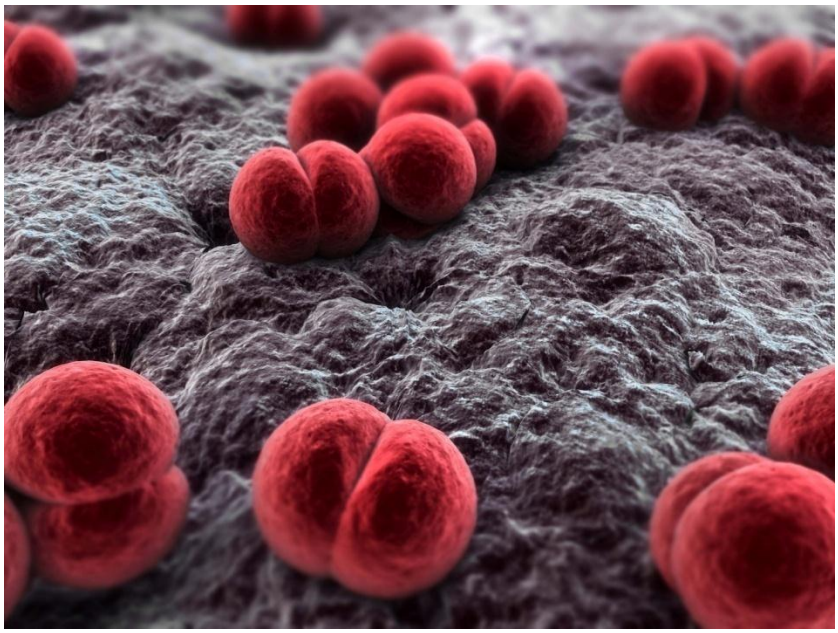


Figura 1: Fotomicrografia eletrônica de corte de meninge humana contendo *N. meninigitidis*. Atentar para a sua forma diplococóide (Figura reproduzida de www.bioquell.com).

N. meningitidis pode colonizar o hospedeiro sem, no entanto, afetá-lo. Este fenômeno é conhecido, do ponto de vista do hospedeiro, como estado de portador (Yasdankhah, 2004). Em situações não-epidêmicas, 10 a 37% da população saudável alberga o

meningococo nas vias aéreas superiores. O estado de portador pode ser crônico (durando meses), intermitente, ou agudo (passageiro). Ocasionalmente, o meningococo pode transpor a barreira mucosa e penetrar na corrente sanguínea (Doulet, 2006) produzindo uma plêiade de patologias que vão desde a meningite purulenta, à artrite séptica, pneumonia, pericardite, sinusite, otite média, conjuntivite e o choque séptico, mais conhecido como meningococcemia ou doença meningocócica. Destas, as mais comuns e também as mais agressivas são a meningite piogênica (purulenta) e a meningococcemia (choque séptico), ambas com alto índice de mortalidade, acometendo principalmente indivíduos adultos jovens (Swarteley, 1997; Taha, 2002; Stephens, 2007).

Os fatores que levam à modificação do estado de portador e ao estabelecimento do estado de doença ainda são pouco compreendidos (Nassif, 2002). Estudos mostram que *N. meningitidis* causa doença sistêmica em uma minoria destes portadores, o que leva a crer que o estabelecimento da doença possa ser um "acidente de percurso" no ciclo biológico da bactéria (Rosenstein, 2001). Na maioria dos casos, o estado de portador constitui um evento imunizador, o qual resulta no aumento da produção, pelo hospedeiro, das três principais classes de anticorpos: G, M, e A (Yasdankhah, 2004; Stephens, 2007).

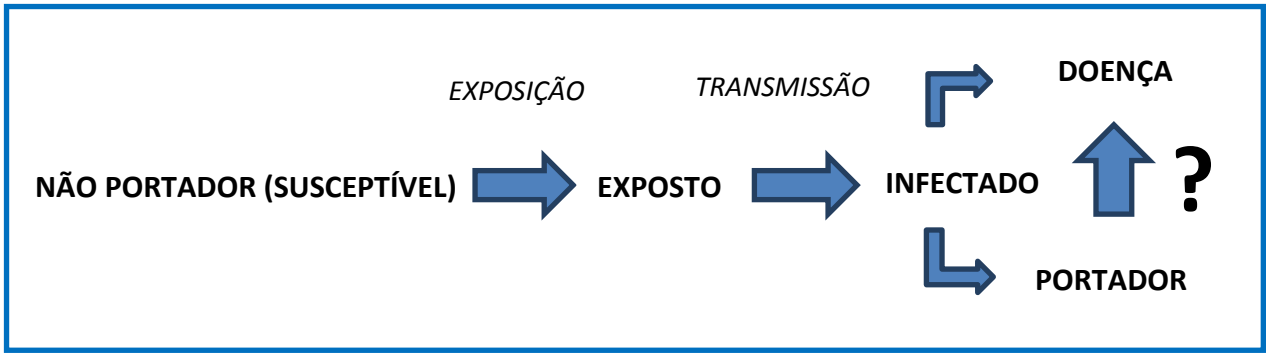


Figura 2. Fluxograma Representativo da Aquisição de *N. meningitidis* e Desenvolvimento de Doença ou Estado de Portador (Adaptado de Yasdankhah, 2004).

As bases genéticas e fisiopatológicas da virulência do meningococo ainda não são bem compreendidas. Dados epidemiológicos e experimentais sugerem que o polissacarídeo capsular, o qual constitui o principal fator de virulência desta bactéria e também um dos fatores de virulência mais bem definidos e estudados até o momento, é necessário, mas não é suficiente para conferir toda a virulência deste microorganismo. (van Dauren, 2000; Nassif, 2002; Stephens, 2007).

Ocasionalmente, o meningococo pode transpor a barreira mucosa e penetrar na corrente sanguínea (Doulet, 2006) produzindo uma plêiade de patologias que vão desde a meningite purulenta, à artrite séptica, pneumonia, pericardite, sinusite, otite média, conjuntivite e o choque séptico, mais conhecido como meningococcemia ou doença meningocócica. Destas, as mais comuns e também as mais agressivas são a meningite piogênica (purulenta) e a meningococcemia (choque séptico), ambas com alto índice de mortalidade, acometendo principalmente indivíduos adultos jovens (Swarteley, 1997; Taha, 2002; Stephens, 2007).

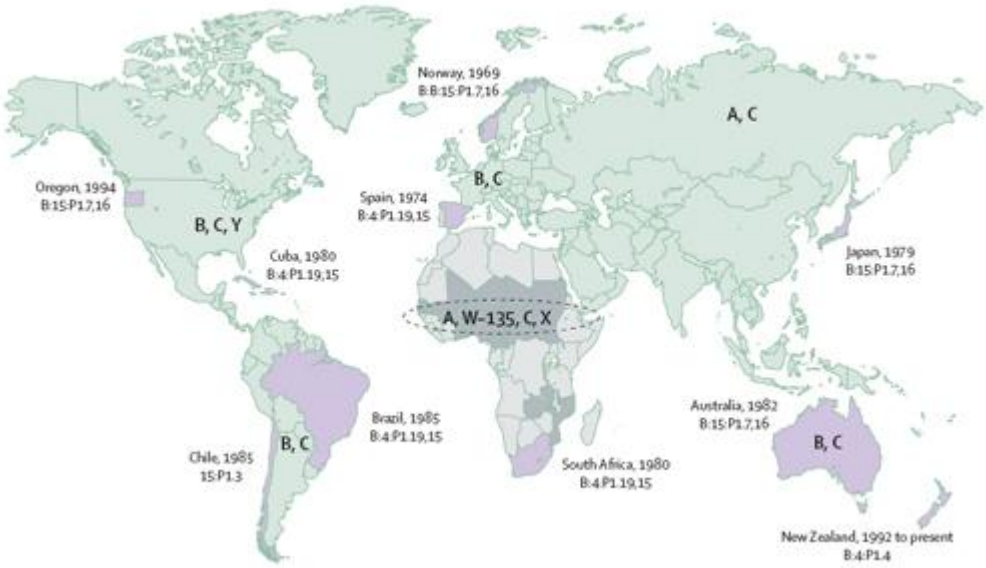
Com base no polissacarídeo capsular é feita a classificação da espécie em 13 diferentes sorogrupos: A, B, C, D, 29E, H, K, I, L, X, Y, W135 e Z, existindo algumas controvérsias quanto ao reconhecimento do sorogrupo D (Hitchcock, 1989; Antignac, 2003; Aspholm , 2010; Hill, 2010) . Destes, cinco sorogrupos, A, B, C, Y e W135, são os mais frequentemente envolvidos em surtos epidêmicos, sendo também os mais associados a casos graves de meningite e doença invasiva (meningococcemia). Os três primeiros sorogrupos são responsáveis por cerca de 90% dos casos, fazendo-se a ressalva de que o sorogrupo A está restrito aos países da África Subsaariana (tais países compõem o chamado "Cinturão Africano da Meningite" (Kwara, 1998)). Desta forma, conclui-se que os sorogrupos B e C

são os principais responsáveis pelos casos de meningite e meningococcemia em todo o mundo, inclusive em países desenvolvidos.(Kwara, 1998; Rosenstein, 2001, Taha, 2002; Yasdankhah, 2004).

Estruturalmente, o polissacarídeo capsular é composto por derivados do ácido siálico ou ácido N-acetil-neuramínico (Dolan-Livengood, 2003), que podem se apresentar como homopolímeros (no caso dos sorogrupos B e C, compostos pelo ácido α2,8 e α2,9 N-acetil neuramínico, respectivamente) ou heteropolímeros (como no caso dos sorogrupos Y e W135, compostos por 6D-glicose e 6D-galactose ligados ao ácido α-1,4-N neuramínico, respectivamente). Já o sorogrupo A tem o seu polissacarídeo capsular composto por α1,6-N-acetil-manosamina-1-fosfato. (Tzeng, 2000). A composição química dos principais sorogrupos está resumida na Tabela 1.

Tabela 1: Composição química da cápsula polissacarídica dos principais sorogrupos de *N. meningitidis* (Tabela adaptada de: Tzeng, 2000; Taha, 2002).

SOROGRUPO	CÁPSULA POLISSACARÍDICA
A	α-*N*-acetil *D*-manosamina -1-fosfato
B	α-2,8, ácido *N*-acetilneuramínico
C	α-2,9, ácido *N*-acetilneuramínico
W135	D-galactose e ácido *N*-acetilneuramínico 1:1
Y	D-glicose e ácido *N*-acetilneuramínico 1:1



A

B

Figura 3A: Distribuição Mundial dos Principais Sorogrupos de *N. meningitidis*. Nota: no mapa, a presença do sorogrupo A na Ásia necessita de maior comprovação científica. Figura 3B: Representação do Cinturão Africano da Meningite. (Figuras reproduzidas de www.who.int - World Health Organization). Meningitis belt = Cinturão da meningite; Atlantic ocean = Oceano Atlântico; Indian ocean = Oceano Índico.

Os sorogrupos são identificados através de técnicas de soroaglutinação, ou empregando-se anticorpos monoclonais (MAbs) em técnicas de *ELISA* (*Enzyme-Linked ImmunoSorbent Assay*) ou *dot-blotting* (Morello*,* 1991). Mais recentemente, as técnicas de biologia molecular, como a Reação em Cadeia da Polimerase (PCR) e o PCR em Tempo Real (qRT-PCR) têm sido amplamente utilizadas. (Taha, 2000; Pollard*,* 2002; Baethgen*,* 2003).

A membrana citoplasmática externa do meningococo contém pili (ou fímbrias), proteínas de membrana externa (OMPs – *outer membrane proteins*), lipooligossacarídeos (LOS) e uma cápsula polissacarídica. Este patógeno apresenta uma variabilidade genética extraordinária que afeta particularmente a natureza antigênica das proteínas de superfície. Tais proteínas são alvos frequentes do fenômeno de transferência horizontal de genes e recombinação homóloga, o que gera a aquisição de novos alelos, com modificação da sequência e estrutura destas proteínas, o que leva a uma maior diversidade antigênica e, consequentemente, a um escape mais eficiente do sistema imune do hospedeiro.(Aguileta, 2009)

Em alguns indivíduos portadores sadios, observou-se a ausência da cápsula polissacarídica em cepas cultivadas à partir da orofaringe, o que leva a crer que além de promover a evasão do sistema imune do hospedeiro (visto que os anticorpos circulantes reconhecem especificamente os antígenos capsulares), a ausência da cápsula possibilita uma melhor adesão da bactéria às células epiteliais do revestimento mucoso da faringe, contribuindo para a perpetuação do estado de portador (Yasdankhah, 2004). A ausência da cápsula, nestas circunstâncias, é explicada pelo fenômeno da variação de fase, ao qual estão sujeitos a maioria dos antígenos de *N. meningitidis.* Este fenômeno é caracterizado pela

expressão seletiva de determinadas proteínas, mediante sensoriamento do meio. Sob o ponto de vista molecular, a bactéria é capaz de alterar a matriz de leitura do gene em questão através de mudanças no número de repetições de nucleotídeos, "ligando" ou "desligando" o gene de acordo com sinais que são enviados do meio externo (Swartley, 1997). A Figura 4 apresenta um desenho esquemático dos principais componentes antigênicos do meningococo.

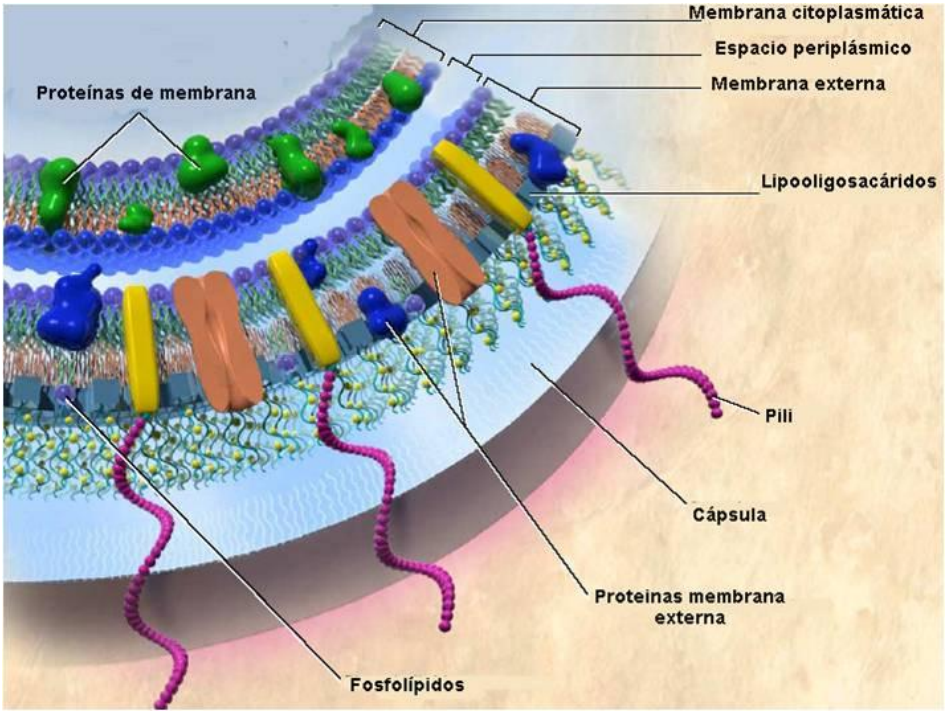


Figura 4. Representação Esquemática da Membrana Externa do Meningococo (Figura reproduzida de The New England Journal of Medicine v.344, n.18, 2001).

1.2 *Haemophilus influenzae*

A bactéria *Haemophilus influenzae* constitui um cocobacilo gram-negativo, não esporulado, e não flagelado (Figura 5). Esta é a espécie mais conhecida do gênero *Haemophilus*, por se tratar de um importante patógeno humano, além de ser o único

membro patogênico deste Gênero. O nome *Haemophilus* deriva da necessidade da presença de sangue no meio de cultura, devidoà incapacidade desta bactéria de sintetizar partes importantes do sistema do citocromo, importante componente da cadeia respiratória. Desta forma, o *H. influenzae* obtém a matéria prima de que necessita à partir da fração heme da hemoglobina sanguínea, também conhecida como fator X (hemina). Além da suplementação com sangue, o meio de cultura também deve ser suplementado com o co-fator nicotinamida adenina dinucleotídeo (NAD), conhecido como fator V (Tortora, 2000; McCrea, 2008).

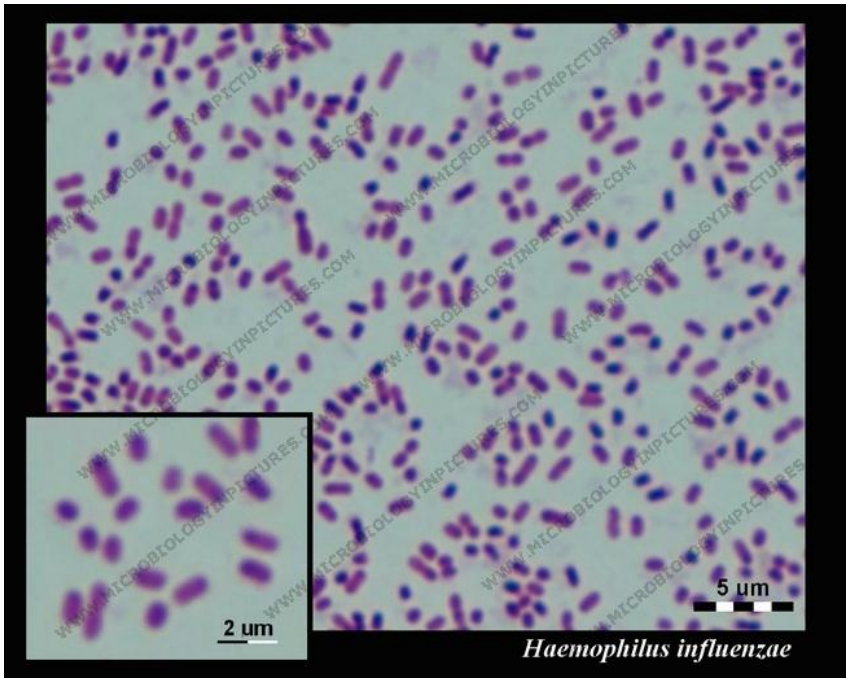


Figura 5: Esfregaço contendo. *H. influenzae* corado pelo método de Gram. Atentar para a sua forma cocobacilífera (Figura reproduzida de www.microbiologyinpictures.com).

Em 1995, esta bactéria foi o primeiro organismo a ter o seu genoma inteiramente sequenciado graças ao tamanho pequeno do seu DNA cromossomal e pelo fato de ser naturalmente competente ao processo de transformação bacteriana.(Fleishmann, 1995)

Bioquimicamente, o *H. influenzae* pode ser subdividido em biotipos e em sorotipos, a depender do métododе tipagem empregado. A divisão em biotipos baseia-se na presença e produção das enzimas ornitina descarboxilase e urease, na fermentação do açúcar D-xilose e na produção de indol. Esta classificação, permite subdividir a espécie em 9 biotipos: I, II, III, IV, V, VI, VII, VIII e o biotipo*aegyptius* (Kilian, 1976).

O *H. influenzae* pode ser dotado ou não de cápsula polissacarídica. Similarmente ao meningococo, a cápsula polissacarídica é considerada o principal antígeno deste microorganismo e sua estrutura antigênica é utilizada para fins de tipagem e classificação das linhagens capsuladas, que são divididas em 6 sorotipos, a – f. (Falla, 1994).

Desta forma, considera-se a existência dos *H. influenzae* não capsulados e, portanto, não tipáveis (NTHi –*Non Typeable Haemophilus influenzae*); dos capsulados, porém não tipáveis, que também são considerados NTHi, e dos capsulados tipáveis, os quais pertencem aos sorotipos a-f (tabela 2)(Falla, 1994; Foxwell, 1998; McCrea, 2008).

Tabela 2: classificação da espécie *H. influenzae* segundo a presença e a tipagem sorológica da cápsula polissacarídica.

PRESENÇA DE CÁPSULA	TIPAGEM	CLASSIFICAÇÃO
SIM	**SIM**	**CAPSULADO (sorotipo a-f)**
SIM	**NÃO**	**NTHi**
NÃO	**NÃO**	**NTHi**

H. influenzae é um patógeno ocasional e constitui um membro comensal da nasofaringe humana (já não mais considerado estrito ao hospedeiro humano) (Marrs, 2001).Por motivos ainda não muito bem compreendidos, esta bactéria pode ultrapassar a barreira mucosa da nasofaringe, causando uma plêiade de patologias do trato respiratório superior e inferior, que vão desde aquelas mais autolimitadas até aquelas mais graves e invasivas. As linhagens não capsuladas geralmente estão associadas a patologias mais autolimitadas, como, por exemplo, a otite média aguda, sinusite, amigdalite, faringite, epiglotite, etc, Também causam pneumonia em indivíduos portadores de fibrose cística pulmonar e em portadores de doença pulmonar obstrutiva crônica (DPOC) (McCrea, 2010). Estas linhagens apresentam-se bastante pleomórficas, quando isoladas do escarro, nas quais se observa longos filamentos. A coloração de gram, quando feita de maneira inapropriada pode, para estes microorganismos, gerar muitos resultados falso-positivos, já que os filamentos tendem a ser fortemente violáceos (Figura 6). Estes resultados falso-positivos, associados à presença de tais filamentos, podem levar muitas vezes a um diagnóstico errôneo de *Streptococcus pneumoniae.*(Foxwell, 1998)

Figura 6: Coloração de gram de um esfregaço de escarro humano mostrando *H. influenzae* não capsulados. A presença de filamentos gera uma coloração de Gram dúbia (Figura reproduzida de textbookofbacteriology.net).

As linhagens capsuladas também colonizam assintomaticamente o trato respiratório superior humano, sendo, porém, mais associadas a doenças invasivas como pneumonia e meningite, acompanhada ou não de septicemia. Principalmente as linhagens capsuladas do sorotipo b estão associadas a meningite purulenta, ocorrendo predominatemente em crianças entre 6 meses a 5 anos de idade (Moxon, 1985; Moxon, 1988).

Até há cerca de duas décadas atrás, o estudo desta espécie voltava-se principalmente para o *H. influenzae* sorotipo b devido a sua agressividade, e à alta taxa de sequelas (Moxon, 1988). Desde a introdução, em 1990, da vacina conjugada para o tipo b, doenças invasivas por este tipo de *Haemophilus*se tornaram mais raras, tendo praticamente se extinguido em países desenvolvidos. Em contrapartida, tem se observado um maior número de relatos de doenças mais graves associadas aos *Haemophilus influenzae* não capsulados (NTHi), antes associados somente a doenças oportunistas ou a casos menos severos de sinusite, bronquite e otite média. (McCrea, 2008; McCrea, 2010).

1.3 Transferência Lateral de Genes em Procariotos

A transferência lateral de genes (TLG), também conhecida como transferência horizontal de genes, constitui um dos exemplos mais marcantes do dinamismo de genomas procariotos. Quando se analisa um genoma bacteriano, constata-se que aproximadamente 18% do seu conteúdo foi adquirido por meio de TLG (Findlay, 2009). A maioria dos genes adquiridos desta maneira foram mantidos nas populações descendentes por seleção natural, por conferir algum tipo de "fitness" ou vantagem adaptativa que permite aos indivíduos explorarem de uma forma mais eficaz os recursos do seu meio, quer seja captando melhor

recursos nutritivos, resistindo a variações extremas de temperatura ou evadindo do sistema imune do hospedeiro, no caso de um comensal. (Freeman, 2007).

A TLG caracteriza-se pela transferência de um segmento de DNA de uma espécie para outra, ou mesmo de um gênero para outro, ocorrendo, portanto, a aquisição de DNA estranho entre genomas distintos (os quais nem sempre são proximamente aparentados). Tal segmento de DNA estranho pode ser aleatório ou pode conter um ou mais genes, ou mesmo um operon inteiro. Este fenômeno, todavia, não se encerra no processo da transferência em si. O DNA exógeno precisa entrar na célula receptora, integrar-se ao seu genoma por recombinação homóloga, e ser fixado nas populações descendentes por seleção natural. Para isto, o gene (ou genes) precisa ser expresso e contribuir com um produto que confira à bactéria uma vantagem seletiva. O termo lateral ou horizontal é utilizado para ressaltar que o compartilhamento de alelos ocorre entre espécies distintas, porém dentro de uma mesma geração. Desta forma, a TLG distingue-se da transmissão vertical de genes, na qual alelos são passados adiante de uma geração para a outra, porém dentro de uma mesma espécie (Freeman, 2007).

A TLG em procariotos pode ocorrer de três maneiras distintas (Figura 7) (Tortora, 2000):

1) Transdução: quando a transferência de material genético cromossomal se dá por meio da infecção da bactéria por um bacteriófago (vírus).

2) Conjugação:quando ocorre a transferência de estruturas extracromossomais (plasmídios) entre duas bactérias (uma doadora e outra receptora) por meio da formação de um pílus sexual.

3) Transformação: a qual será abordada no próximo item.

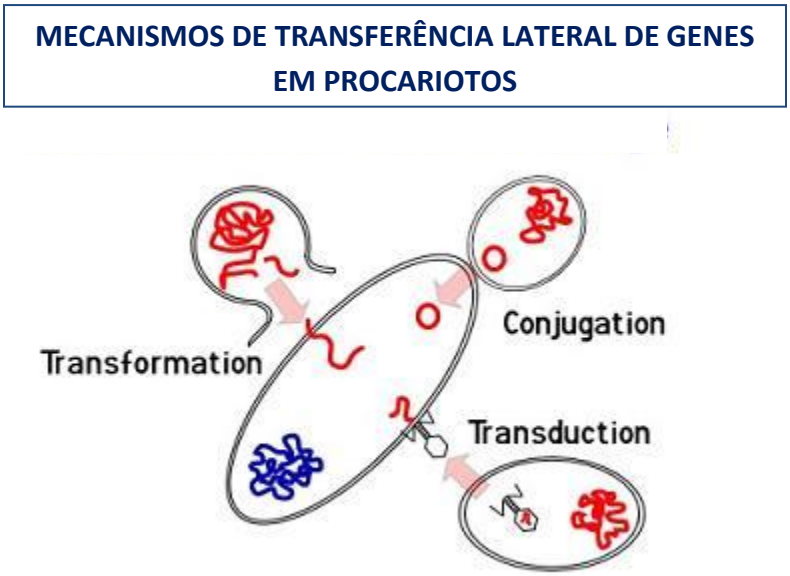


Figura 7: Representação esquemática dos mecanismos de Transferência Lateral de Genes (Figura reproduzida de science.kennesaw.edu). Transformation = Transformação; Conjugation = Conjugação; Transduction = Transdução.

A constatação e a confirmação de que a presença de um determinado gene ao longo de um genoma se deve a um evento de TLG pode ser difícil. Uma das primeiras estratégias é verificar se o DNA em questão confere alguma função inteiramente nova e até então não reconhecida naquela bactéria. Outra estratégia consiste em documentar a frequência de pares guanosina-citosina (GC) relativa à de pares adenina-timina (AT) no segmento sob investigação, comparando-a com a mesma proporção em relação ao resto de genoma. Eventos recentes de TLG costumam exibir proporções diferentes de pares GC em relação ao do genoma receptor.No entanto, quando o evento é muito antigo na escala de tempo ou o genoma da bactéria receptora está sob um regime de pressão seletiva muito forte, as características inerentes àquele segmento podem ser apagadas, de forma que a detecção da

TLG torna-se mais difícil, ou mesmo impossível. Outra estratégia válida é observar a presença, no segmento em questão, de códons raros ou não usuais para o genoma analisado. A varredura do genoma à procura de segmentos idênticos aos de transposases, ou à procura de sequências de inserção de vírus ou de sequências "uptake" é outro modo de se inferir se um determinado alelo originou-se por TLG (Freeman, 2007).

No entanto, o método clássico e mais elucidativo de se identificar e confirmar uma TLG é construir uma árvore filogenética para as espécies envolvidas no processo (topologia esperada) e uma árvore para o gene em questão (topologia observada) (Figura 8). Ao se comparar as topologias das duas árvores, será observado que o ramo contendo o gene sob investigação colocará a espécie uma posição anômala em relação à topologia esperada (ou seja, em relação à topologia da árvore de espécies). Realizando-se um teste de confiabilidade amostral, se a nota do ramo "anômalo" na árvore de genes for igual ou superior ao do mesmo ramo na árvore de espécies, isto indica a ocorrência da transferência lateral do gene em questão. A evidência mais convincente de todas, porém, ocorre quando todas as estratégias previamente citadas somadas às análises filogenéticas são concordantes com a hipótese de transferência lateral do gene (Freeman, 2007).

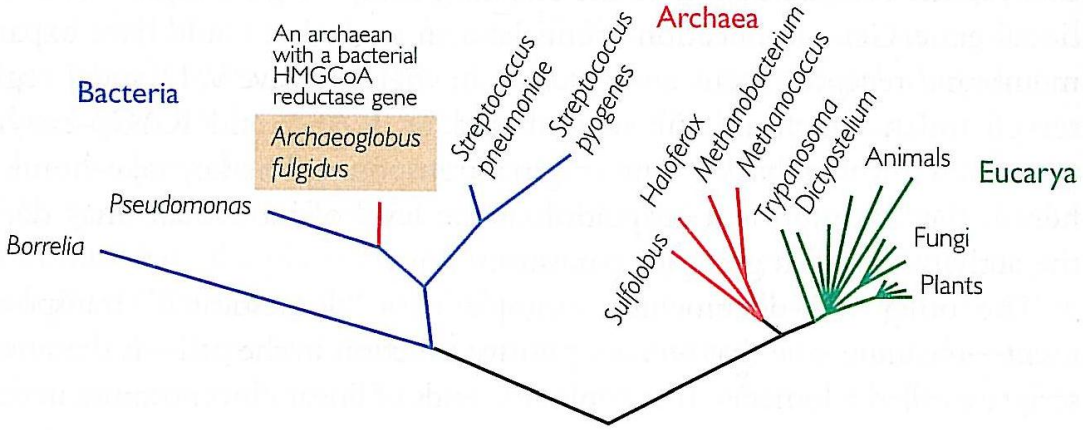


Figura 8: Árvore filogenética evidenciando um ramo anômalo. *Archaeoglobus fulgidus* é um membro do domínio *Archea* que recebeu o gene *HMGCoA reductase* de um membro do domínio *Bacteria*. Uma árvore filogenética para este gene revela a sua posição anômala em relação ao esperado (Figura reproduzida de Freeman, 2007). Animals = Animais; Fungi = Fungos; Plants = Plantas.

Bactérias que sofrem freqüentes de eventos de TLG exibem um genoma extremamente mosaico (Findlay, 2009). Desta maneira, o processo de TLG pode ser compreendido como um evento promotor da microevolução bacteriana na medida em que a captação e integração de um segmento inteiramente novo de DNA, além de aumentar a variabilidade genética do microorganismo, confere a uma ampla gama de linhagens receptoras um fenótipo vantajoso em relação ao seu meio, como novas capacidades metabólicas e virulentas, permitindo-lhes explorar novos recursos alimentares e/ou ocupar novos habitats, ou mesmo novos nichos ecológicos. Portanto, a TLG tem uma implicação preponderante na coevolução patógeno X hospedeiro e à nível da Saúde Pública. Sob a ótica da coevolução patógeno X hospedeiro, a rápida aquisição de um novo fator de virulência por uma dada população de bactérias permite o escape do sensoreamento do sistema imune e o estabelecimento do estado de doença. Do ponto de vista da Saúde Pública, isto explica os freqüentes fenômenos de escapes vacinais, os inúmeros casos de

resistência crescente a antimicrobianos e o constante surgimento de "superbactérias" (Freeman, 2007).

Vale ressaltar que o genoma "central" bacteriano (genes envolvidos em processamento de informação como replicação do DNA, transcrição e tradução) raramente é afetado por eventos de TLG (Findlay, 2009).

1.4 O Estado de Competência Natural ao Processo de Transformação Bacteriana

A competência natural ao processo de transformação bacteriana é um estado fisiológico, geneticamente programado, que permite a eficiente captação de DNA livre no meio externo. A transformação natural distingue-se da transformação induzida artificialmente, na qual métodos físico-químicos induzem a despolarização da membrana celular e a abertura de canais iônicos. O estado de competência natural é um processo complexo, altamente regulado, o qual tem sido alvo de intensas pesquisas. (Redfield, 1993; Redfield, 1997; Dubnau, 1999).

DNA livre no meio externo é encontrado em abundância e pode ter duas origens. A mais comum delas é a lise de outras bactérias, liberando fragmentos de DNA dupla fita no meio (Dubnau, 1999; Chen, 2004). Tem sido observado também que algumas bactérias, como *N. gonorrhoeae*, tem a capacidade de secretar ativamente DNA no meio externo através do sistema de secreção tipo IV (Steinmoen, 2002; Steinmoen, 2003).

O transporte de DNA do meio externo até o interior do citoplasma requer a participação ordenada de milhares de proteínas. Verificou-se que as proteínas que tomam

parte neste processo são as mesmas, ou bastante similares, àquelas envolvidas na montagem do pilus tipo IV (Figura 9B) e no sistema de secreção bacteriana tipo II (Figura 9A), especulando-se se haveria um pilus bacteriano modificado para fins de captação de DNA exógeno (Chen, 2004). Este complexo protéico está associado, no lado interno da membrana citoplasmática, a um outro complexo onde se encontra uma translocase de DNA (Chen, 2004).

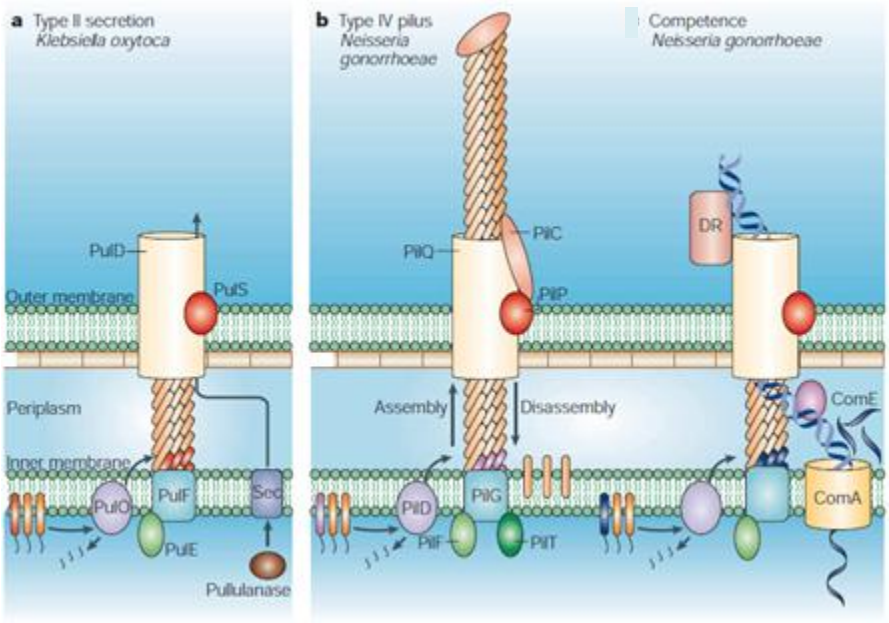


Figura 9A: Representação esquemática do sistema de secreção bacteriana do tipoII em *Klebsiella oxytoca*. Type II secretion = Sistema de secreção tipo II; Outer Membrane = Membrana externa; Periplasm = Periplasma; Inner Membrane = Membrana interna. Figura 9B: Comparação do sistema de secreção tipoIV e do complexo protéico associado ao estado natural de competência em *Neisseria gonorrhoeae* (Figura reproduzida de Dubnau, 1999). Type IV pilus = Pilus tipo IV; Assembly = Montagem; Disassembly = desmontagem; Competence = Competência.

A competência natural não é um estado fisiológico permanente. Sabe-se que ele é induzido em ocasiões específicas, sendo regulado por sinais externos tais como *quorum sensing* e sinais nutricionais (MacFadyen, 2001; Claverys, 2002; Hamoen, 2003). Portanto,

é possível que muito mais espécies do que as reconhecidas atualmente exibam a competência natural ao processo de transformação, no entanto as condições nas quais elas desenvolvem tal estado ainda são pouco compreendidas (Claverys, 2003).

Tanto bactérias gram-positivas quando gram-negativas exibem o estado de competência natural. O transporte de DNA exógeno pela membrana citoplasmática interna é similar em ambas, e algumas proteínas implicadas neste processo são ortólogas. As etapas iniciais, no entanto, diferem devido à presença de uma membrana externa em gram-negativos e de uma espessa parede celular em gram-positivos (Dubnau, 1999) (Figura 10). O mecanismo de transformação em gram-positivos, por não ser alvo deste estudo, não será abordado. A nomenclatura utilizada será aquela descrita para *Neisseria gonorrhoeae*, mas o mecanismo é similar para as bactérias *H. influenzae* e *N. meningitidis* (Redfield, 1993; Redfield 1997).

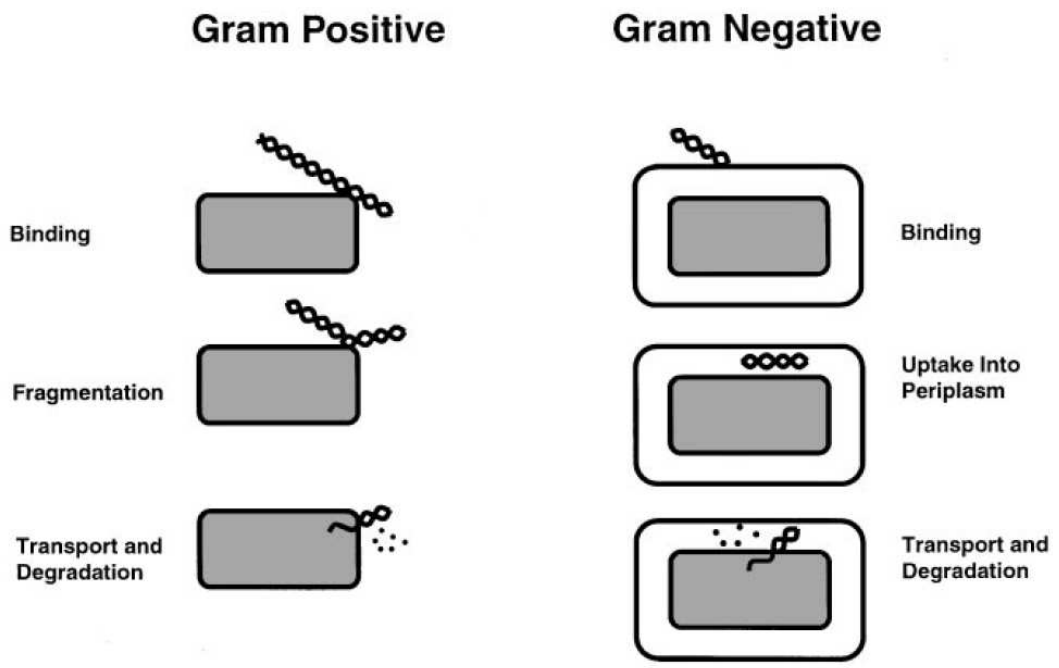


Figura 10: Diferenças nas etapas de transformação em bactérias gram-positivas (lado esquerdo) e gram-negativas (lado direito) (Figura reproduzida de Dubnau, 1999). Gram Positive = Gram Positivo; Gram Negative = Gram Negativo; Binding = Ligação; Fragmentation = Fragmentação; Uptake into the periplasm = Captação para dentro do periplasma; Transport and Degradation = Transporte e Degradação.

Em bactérias gram-negativas, o segmento de DNA exógeno, além de atravessar a parede celular e a membrana celular interna, deve atravessar também a membrana celular externa. Portanto, etapas adicionais estão envolvidas no sistema de transformação em gram-negativos. Além disso, a etapa de interação inicial do DNA com a superfície celular em bactérias gram-negativas também difere em relação às bactérias gram-positivas (Dubnau, 1999).

Primeiramente, o DNA na forma de dupla fita liga-se irreversivelmente à superfície celular através da sua extremidade 3'. Esta ligação, no entanto não é indiscriminada. Ela é dependente da presença de sequências ou motivos curtos denominados de sequências

uptake (US) flanqueando o segmento de DNA a ser incorporado. São essas USs que são reconhecidas na superfície celular, sendo específicas para cada espécie bacteriana. Inicialmente pensou-se que a presença de USs tornava a etapa de ligação de DNA exógeno espécie-específico, porém, posteriormente, foi constatado que, mesmo em presença das USs, havia a troca de DNA entre espécies e mesmo entre gêneros distintos (Kroll, 1998; Findlay, 2009). Geralmente não ocorre fragmentação do segmento de DNA que está sendo incorporado, como ocorre em gram-positivos (Biswas, 1986).

A montagem do complexo protéico associado ao estado de competência natural resume-se da seguinte maneira (Figura 11): PilC, uma pilina (pois é necessária para a formação do pilus tipo IV) e também uma proteína periférica de membrana externa, juntamente com PilQ, uma secretina (um complexo protéico transmembrana composto de várias subunidades idênticas, formando um canal aquoso na membrana celular externa) são necessárias para o transporte do DNA na forma de dupla fita pela membrana celular externa (Fascius, 1996). PilE, outra pilina, polimeriza-se formando um canal que liga a membrana celular externa à interna, atravessando o espaço periplasmático e a parede celular. Associadas externamente a PilE estão as proteínas Tpc e ComL, hidrolases que catalizam a hidrólise do ácido murâmico, componente da parede celular bacteriana, facilitando a formação do canal aquoso (Fascius, 1996). Uma proteína receptora de DNA dupla fita localiza-se no interior deste canal aquoso, bem como uma nuclease responsável pela transformação do DNA dupla fita em simples fita, bem como pela lise de uma das fitas. Um complexo protéico na forma de poro, situado na membrana interna e formado por várias subunidades da proteína ComA, permite a passagem de DNA simples fita para o citoplasma

(Fascius, 1993). O transporte do DNA para o citoplasma é facilitado por uma proteína transportadora associada a ComA.

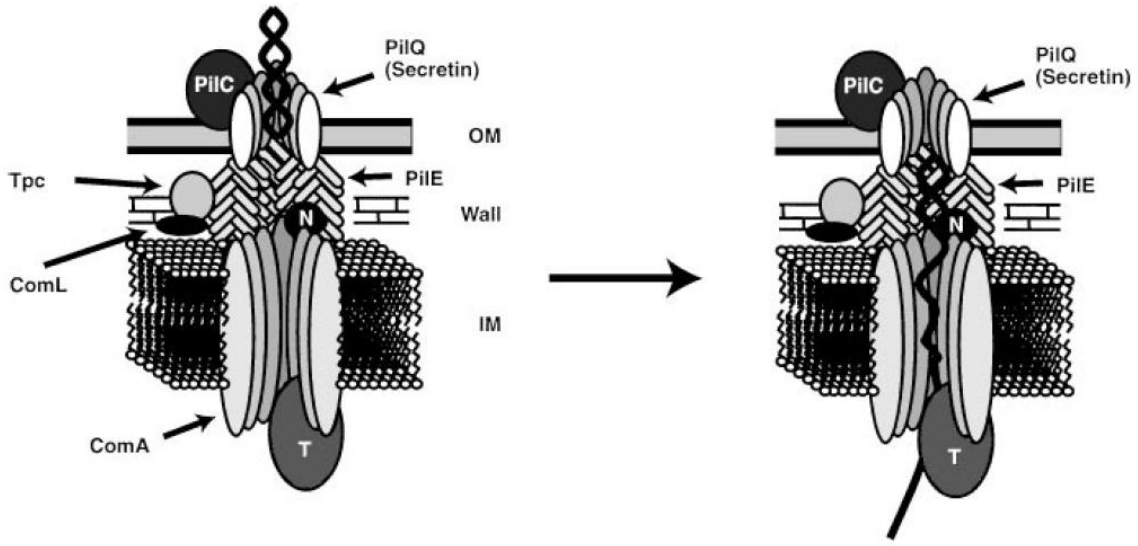


Figura 11: Representação esquemática do complexo protéico (provável pilus modificado) responsável pela transformação em bactérias gram-negativas (Figura reproduzida de Dubnau, 1999). Secretin = Secretina; OM = Outer membrane; Wall = Parede; IM = Inner Membrane

Uma vez no interior do citoplasma, o DNA simples fita é envolto por um conjunto de pequenas proteínas globulares que impede a sua degradação por nucleases intracitoplasmáticas. A este complexo DNA-proteína denomina-se transformassomo. Estas proteínas também são responsáveis por conduzir o DNA até o cromossomo bacteriano, onde ocorre a recombinação homóloga do DNA (Dubnau, 1999)

Todas as etapas descritas anteriormente são mediadas por proteínas denominas PSTC. Estas proteínas são necessárias para a formação do pilus tipo IV, para a montagem do sistema de secreção tipo II e para a motilidade, por isso o seu nome: PSTC (**p**ilus/**s**ecretion/**t**witching motility/**c**ompetence). Por serem amplamente distribuídas e

conservadas tanto em gram-negativos quanto em gram-positivos, sugeriu-se que as proteínas PSTC cumpriam algum tipo de mecanismo ancestral, anterior à divergência entre gram-positivos e gram-negativos (Dubnau, 1999).

A capacidade de adquirir DNA exógeno por transformação natural é um recurso amplamente difundido entre bactérias, reconhecendo-se atualmente mais de 40 espécies dotadas de tal peculiaridade, as quais estão distribuídas entre os mais variados grupos taxonômicos (Lorenz, 1994).

Este é um processo altamente regulado, visto que requer a participação ordenada de dezenas de proteínas e deve estar em fina sintonia com as necessidades de cada microorganismo . Especula-se acerca de qual força evolutiva molda e mantém um sistema tão elaborado em tantas espécies bacterianas. Várias hipóteses foram formuladas, sendo as principais: 1) DNA exógeno como fonte nutritiva; 2) DNA como substrato para o mecanismo de reparo e 3) DNA como fonte de diversidade genética (Dubnau, 1999).

- DNA como fonte nutritiva:

Foi proposto que o estado de competência natural evoluiu para fins de captação de DNA exógeno como suplemento nutritivo (ao ser internalizado no citoplasma, o DNA seria fragmentado e seus nucleotídeos integrados nas vias de síntese metabólica) (Redfield, 1993; Redfield, 1997). Muitos são os argumentos contra esta hipótese. Um deles é o fato de algumas bactérias possuírem mecanismos mais simples para suprir tal finalidade, como a secreção de nucleases no meio externo e sistemas mais simples de captação de nucleotídeos (Dubnau, 1999). Um outro argumento parece bastante robusto: a maquinaria de transformação é bastante elaborada e dispendiosa (em termos de consumo de energia), no

entanto DNA é incorporado na célula na forma de fita simples, portanto metade do potencial nutritivo seria simplesmente descartado no meio. Isto, seria uma aparente contradição, já que manutenção de um mecanismo tão requintado, não deveria pressupor tanto desperdício (Dubnau, 1999). Um outro argumento contra é que, em gram-negativos, o mecanismo de captação exibe especificidade, não sugerindo, desta forma, uma finalidade nutritiva específica. Do contrário, DNA seria captado indiscriminadamente do meio. No entanto, não se pode descartar que , em algum momento da história evolutiva destes microorganismos, a transformação natural tenha preenchido este fim (Dubnau, 1999).

- DNA como substrato para o mecanismo de reparo

Foi proposto que nas bactérias *B. subtilis* e *H. influenzae* a maquinaria de reparo é induzida como parte do regulon da competência natural (Love, 1985; Haijema, 1996). Portanto, a indução do sistema de reparo antes do aparecimento do dano no DNA conferiria uma vantagem seletiva. Por exemplo: quando a bactéria se aproxima da fase estacionária do crescimento, quando o seu metabolismo se torna mais aeróbico, e dano oxidativo ao DNA é mais provável de ocorrer, uma indução preventiva do sistema de reparo evitaria o aparecimento de danos mais sérios no DNA (Dubnau, 1999). No entanto, esta teoria tem sido seriamente criticada pelo fato de que agentes mutagênicos não induzem o estado de competência natural (Redfield, 1993).

- DNA como fonte de variabilidade genética

A terceira, mais popular, e mais bem aceita das hipóteses propõe que a trans formação é um mecanismo introdutor e mantenedor de diversidade no genoma bacteriano (Figura 12). Desta forma, observar-se-ia, numa população, uma maior distribuição de

valores de aptidão, permitindo aos indivíduos portadores de alelos vantajosos ao seu meio, numa dada circunstância, ocupar picos adaptativos, enquanto que aqueles portadores de alelos menos vantajosos experimentariam vales adaptativos. Isto significa dizer que, numa paisagem (ou habitat) em constantes modificações, pouca variação no genoma pode levar seriamente todos os indivíduos de uma população a experimentarem vales adaptativos, o que poderia conduzir toda a população à extinção (Dubnau, 1999).

Toda diversidade genética deriva da ocorrência de mutações no genoma. Todo evento de transformação natural é seguido de recombinação homóloga do novo seguimento de DNA no genoma receptor. Portanto, a recombinação homóloga contribui para o aumento da diversidade genética de uma população de duas maneiras: 1) enquanto evento mutacional, a recombinação constitui um mecanismo gerador de novos alelos na população, prevenindo a redução da diversidade causada por varreduras seletivas que se seguem à fixação de um alelo muito vantajoso numa população numa dada circunstância; 2) através do processo de transformação, novos genes são introduzidos num genoma, conferindo novos recursos e novas capacidades metabólicas ao organismo receptor. (Freeman, 2007).

Visto que a troca gênica por transformação é um evento freqüente em muitas espécies bacterianas, é tentador concluir que a manutenção da diversidade genética é a força evolutiva que explica o surgimento e manutenção do estado de competência natural. No entanto, esta conclusão é apenas especulativa e o debate concernente ao papel evolutivo da transformação natural continua. É digno de nota que os sinais que induzem o estado de competência são específicos para cada espécie. Isto pode significar que o processo de transformação natural pode servir a uma gama de variedades e que cada espécie “aprendeu”

a usar a habilidade de captar DNA exógeno para suprir suas próprias necessidades (Dubnau, 1999).

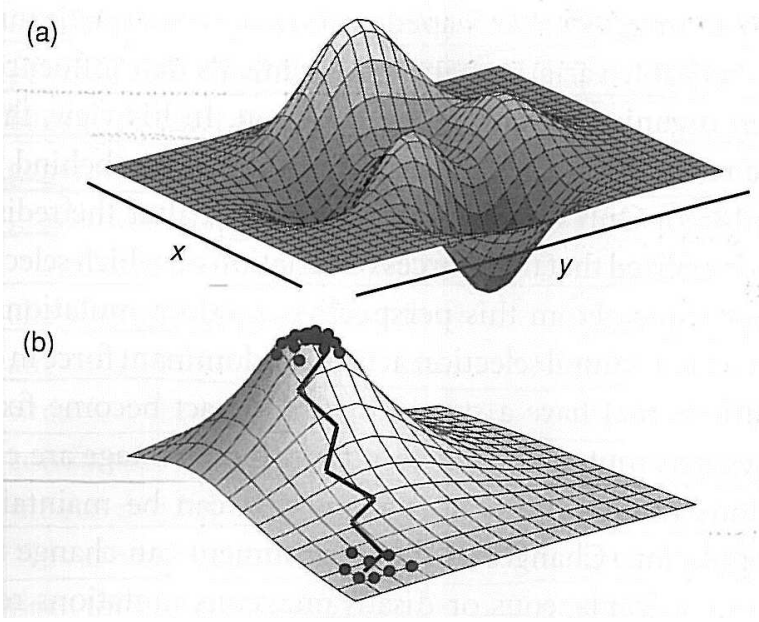


Figura 12A: Representação de uma paisagem em constantes modificações, com picos e vales adaptativos. Figura 12B: Os pontos indicam indivíduos de uma população ocupando picos e vales adaptativos. Aqueles que estão no pico possuem alelos mais vantajosos em relação ao meio (Figura adaptada de Lemey, 2009).

1.5 A Enzima [Cu, Zn] Superóxido Dismutase

A produção de espécies reativas intermediárias do oxigênio (ROS) constitui uma característica intrínseca do metabolismo aeróbico normal, podendo, no entanto, ser deletéria em organismos vivos devido à oxidação de macromoléculas como DNA, lipídeos e proteínas. A oxidação de macromoléculas pode levar, por exemplo, a mutações gênicas, desnaturação e agregação de proteínas, e peroxidação lipídica (devido à reação de ROS com radicais livres derivados do nitrogênio – RNS- também produzidos nas mesmas

circunstâncias, formando um composto altamente tóxico, o perinitrito), com o conseqüente comprometimento de membranas biológicas, o que pode resultar em dano celular generalizado, seguido de morte celular por degeneração ou apoptose (Medina, 2010). Os ROS são constituídos por várias espécies com diferente reatividade química frente a biomoléculas e difusão em meio biológico. Uma característica em comum, no entanto, é a de exibirem alta reatividade e baixa especificidade (Medina, 2010).

Em função da toxicidade associada a estas espécies e para garantir a integridade de organismos vivos, ocorreu o surgimento de um complexo sistema antioxidante intracelular, composto por enzimas específicas, moléculas de baixo peso molecular e sistemas de reparo a biomoléculas. O intricado balanço entre a produção de ROS e a sua eliminação pelo sistema antioxidante garante a homeostasia dos organismos vivos. A situação em que ocorre a produção excessiva de ROS e/ou deficiência na sua eliminação pelo sistema antioxidante denomina-se estresse oxidativo, e pode levar a injúria celular e tecidual irreversível (Medina, 2010).

Superóxido dismutases são metaloenzimas, (enzimas que possuem íons metálicos ligados direta ou indiretamente à sua cadeia polipeptídica), que compõem o sistema antioxidante intracelular. Elas são responsáveis por catalizar a conversão de ânions superóxido em peróxido de hidrogênio e oxigênio molecular, atuando nas etapas iniciais de uma série de reações citoprotetoras, e impedindo a formação do peroxinitrito e do radical hidroxila (McCord &Fridovich, 1969; Fridovich, 1986; Hassan, 1989) (Figura 13).

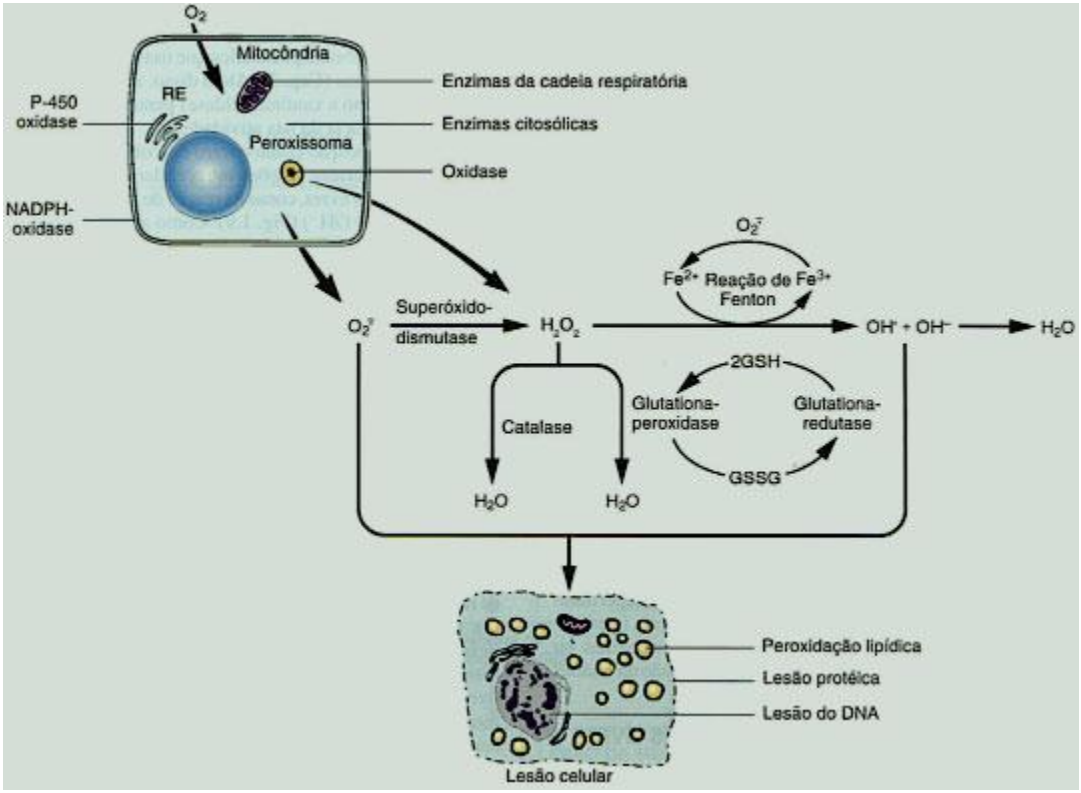
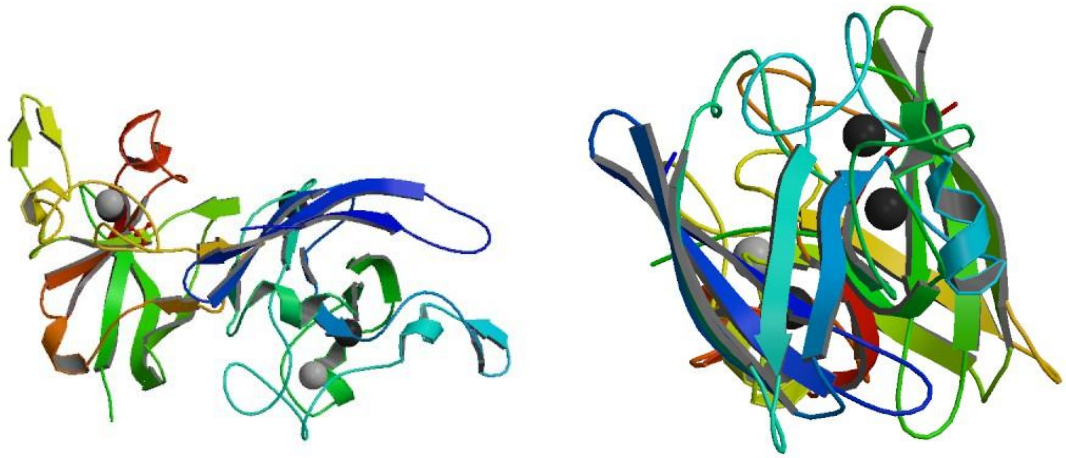


Figura 13: Representação esquemática da atuação da Sod nas etapas iniciais de reações citoprotetoras (Reproduzido de www.sistemanervoso.com).

Estas enzimas possuem distribuição quase universal e, virtualmente, quase todas as células possuem uma ou mais formas da enzima no seu citoplasma. Eucariotos superiores possuem ainda uma forma extracelular (Kroll, 1991). Três formas desta enzima são comumente conhecidas, e sua classificação é feita de acordo com o tipo de íon cofator que se liga ao sítio ativo: SodA tem como cofator o manganês (Mn) e é amplamente encontrada em bactérias e mitocôndrias, na matriz mitocondrial (Grace, 1990), SodB tem como cofator o Ferro (Fe) e é encontrada em bactérias, eucariotos primitivos e em algumas plantas verdes (Van Camp, 1990), e, finalmente, a SodC, que tem por cofatores os íons cobre (Cu) e zinco (Zn), é encontrada no periplasma de bactérias gram-negativas e também no citosol de eucariotos (Bannister, 1987; Grace, 1990).

As Sods A e B são muito similares tanto em termos de estrutura quanto em termos de sequência primária, o que sugere que ambas evoluíram à partir de um ancestral comum (Parker, 1988; Grace, 1990). Em contraste, a SodC bacteriana não apresenta nenhuma semelhança em termos de sequência ou estrutura com as demais Sods procarióticas nem com a SodC eucariótica (esta é um tetrâmero e a SodC bacteriana, um dímero composto por duas subunidades idênticas (Kroll, 1995)) (Figura 14A e B). Desta forma, parece, de fato, que a SodC procariótica evoluíu por um caminho distinto das demais (Kroll, 1991), talvez refletindo uma evolução convergente sobre um fenótipo importante (Grace, 1990; Kroll, 1995).



A **B**

Figura 14A: Estrutura cristalográfica da SodC de *N. meningitidis,* compostas por dois monômeros associados por pontes dissulfeto intramoleculares. Figura 14B: Estrutura cristalográfica de um dos quatro monômeros da SodC bovina (Figuras reproduzidas de www.rcsb.org/pdb)

À princípio, a SodC foi encontrada somente em eucariotos. Apenas em 1974 (Puget, 1974), a primeira SodC bacteriana foi identificada e isolada à partir da bactéria endossimbionte *Photobacterium leiognathi*, o que levou os autores a concluírem

equivocadamente que a SodC bacteriana havia originado-se por transferência lateral de genes à partir do hospedeiro (Puget, 1974; Flohe, 1984; Bannister, 1985). À medida que outras SodsC começaram a ser identificadas e isoladas de outras bactérias, inclusive não simbiontes (*Caulobacter crescentus* (Steinman, 1982; Steinman, 1990), *Paracoccus denitrificans* (Vignais, 1982), *Pseudomonas maltophilia* e *Pseudomonas aeruginosa* (Steinman, 1985), *e Brucella abortus* (Sriranganathan, 1991)), esta proposição caiu por terra, constatando-se que esta forma da enzima é muito comum em bactérias (Kroll, 1995; Kroll, 2003; McCrea, 2008; McCrea 2010).

Com base na comparação entre as seqüências protéicas das formas eucariótica e procariótica, determinou-se quais resíduos de aminoácidos são invariantes e, portanto, cruciais para a atividade da enzima (Kroll, 1995):. Ao todo, 23 dos 186 resíduos de aminoácidos são invariantes: 15 estão nas vizinhanças do sítio ativo, 4 na interface entre os monômeros e 4 compõem a estrutura em barril β (Kroll, 1991).

A estrutura tridimensional da SodC bovina é conhecida em detalhe e os resíduos de aminoácidos críticos para a atividade catalítica já foram identificados (Kroll, 1991). Os íons cofatores, importantes tanto para a atividade enzimática quanto para o correto dobramento da enzima, ligam-se a seis histidinas (HIS 80, HIS 82, HIS 105, HIS 114, HIS 123, HIS 161) e a um resíduo de aspartato (ASP 126). Esses resíduos são altamente conservados também entre as Sods C bacterianas, bem como um resíduo de arginina (ARG 180) na entrada do sítio catalítico, além de dois resíduos de cisteína (CIS 87 E CIS 183) envolvidos no estabelecimento de pontes intramoleculares (Kroll, 1991). Cinco das seis histidinas e todos os outros quatro resíduos conservados são encontrados na sequência da SodC de *H. influenzae,* com exceção da Histidina na posição 82, aquela que liga o cobre. Em *H.*

influenzae, este resíduo foi subtituído por uma tirosina como resultado de uma transição CAT-TAT. A sequência também possui um domínio de 22 AA hidrofóbicos na porção N-terminal, característico de um peptídeo sinal, o que sugere que a enzima é exportada (Kroll, 1991).

Algumas evidências experimentais sugerem que esta enzima é secretada para o periplasma (Steinman, 1987; Steinman 1990). Além disso, a presença de uma região hidrofóbica composta por 22 aminoácidos na porção N terminal, característica de um peptídeo sinal, é um forte indício de que a enzima, de fato, localiza-se fora do citoplasma (Kroll, 1991).

Por localizar-se no periplasma de bactérias gram-negativas (Steinman, 1987; Steinman 1990; Stable, 1994) (Figura 15), e visto que os ROS gerados no interior do citoplasma não podem atravessar a membrana citoplasmática interna (Hassan & Fridovich, 1979) especulou-se que a SodC desempenharia um papel distinto das demais Sods bacterianas, e mais análogo ao da forma extracelular da SodC eucariótica (dismutando ânions superóxido gerados exogenamente (Kroll, 1995)).

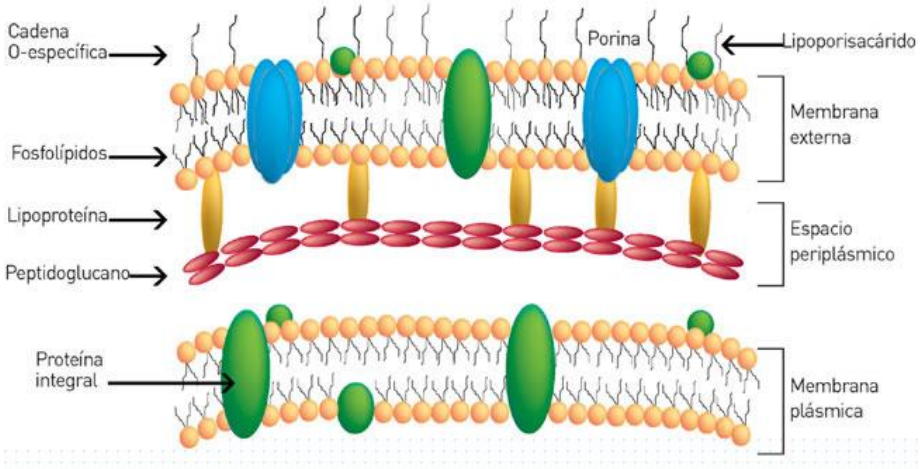


Figura 15: Representação esquemática do espaço periplasmático de bactérias gram-negativas, localização subcelular da enzima SodC. Atentar para o fato de que a membrana plasmática interna impede a passagem de ROS do citoplasma para o periplasma (Figura reproduzida de www.unb.br). Cadena O-específica = Antígeno O; Lipoprisacárido = Lipopolissacarídeo; Espácio periplásmico = Espaço periplasmático.

Em 1970, foi conduzido um ensaio de oxido-redução *in vitro* utilizando-se o sistema da xantina/hipoxantina. Neste estudo foi demostrado que a enzima SodC era capaz de proteger a bactéria *N. meningitidis* da ação de radicais livres gerados no meio extracelular à partir da oxidação da xantina pela enzima xantina oxidase (Fridovich, 1970). Posteriormente diversos ensaios de fagocitose *in vitro* foram conduzidos, comparando-se a performance de bactérias selvagens com a de mutantes deficientes em SodC frente à fagocitose. Estes estudos demonstraram que a enzima SodC contribuía significativamente para escape bacteriano do fagolisossomo (bactérias selvagens eram recuperadas do meio intracelular, enquanto que as mutantes atenuadas morriam), desta forma contribuindo para a virulência do patógeno (Dunn, 2003).

Sabe-se que, durante o "burst" oxidativo (fenômeno que corresponde a uma explosão respitatória em fagócitos do sistema imune quando ativados), macrófagos e neutrófilos

liberam grandes quantidades de radicais livres nas imediações da membrana citoplasmática, o que facilita e estimula a fagocitose anticorpo-mediada, (Dunn, 2003). Radicais livres também são secretados ativamente no interior de fagolisossomos. Grande debate ainda existe, no entanto, acerca de em qual estágio propriamente dito da fagocitose a SodC atua: se num estágio precoce, modulando a associação da bactéria com a superfície celular, se impedindo o engolfamento propriamente dito, ou se tardiamente, promovendo o escape do fagolisossomo (Dunn, 2003; Lo, 2009).

Em 2003, foi conduzido um ensaio de fagocitose *in vitro* utilizando-se monócitos/macrófagos humanos no intuito de comparar o perfil de meningococos selvagens do sorogrupo B com mutantes atenuados para a SodC (Dunn, 2003). Poucas bactérias selvagens foram recuperadas do meio interno, ao passo que grande quantidade foi recuperada do meio externo após 1h de fagocitose. Portanto, foi apontado que, a SodC ao dismutar ânions superóxido gerados na superfície do macrófago dificulta o engolfamento em si,pois torna a bactéria menos vulnerável, mesmo opsonizada (Dunn, 2003).

1.6 A Transferência Lateral do Gene *sodC*

Exemplos marcantes de sequências repetitivas ocorrem no genoma de algumas bactérias naturalmente competentes ao processo de transformação. Estas sequências ocorrem aos milhares e constituem cópias de motivos curtos denominados *DNA Uptake Signal Sequences* ou simplesmente *Uptake Sequences* (US), presentes no genoma de bactérias pertencentes à família *Pasteurellaceae* e também do gênero *Neisseria*. A maioria

destas sequências pode ocorrer flanqueando segmentos aleatórios de DNA ou flanqueando um gene, genes, ou mesmo operons inteiros. (Findlay, 2009).

Apesar da maioria das bactérias naturalmente competentes captarem fragmentos de DNA de dupla fita do meio indiscriminadamente, os membros da família *Pasteurellaceae* e do gênero *Neisseria* captam e integram ao seu genoma apenas aqueles flanqueados pelas suas USs. No gênero *Haemophilus,* a US caracteriza-se por um oligonucleotídio de 29 pb (5'-aAAGTGCGGTnRWWWWWnnnnnnRWWWWW-3') , contendo um domínio central altamente conservado de 9 pb (AAGTGCGGT). Neste caso, "a" representa preferencialmente uma adenina, "n" representa qualquer base, "R" constitui uma purina e "W" pode ser uma adenina ou uma timina. Em *Neisseria,* o motivo é constituído de 10 pb (5'- GCCGTCTGAA -3'). Cada motivo ou US é altamente representada ao longo do genoma do seu próprio organismo (constituindo aproximadamente 17% do mesmo) (Findlay, 2009).

Desta forma, pensava-se, inicialmente, haver uma preferência da espécie pelo seu próprio DNA contra fragmentos de DNA estranho, visto haver diferenças de tamanho e composição nestas USs e porque, até então, não se tinha notícias da transferência de DNA entre bactérias distantemente relacionadas (*H. influenzae* pertence à Classe das δ-Proteobactérias e *N. meningitidis* pertence à Classe das β-Proteobactèrias) (Tortora, 2000) (Figura 16). Agora sabe-se que esta preferência se estende a espécies correlatas (Findlay, 2009) e mesmo entre gêneros distantes como é o caso dos gêneros *Haemophilus* e *Neisseria* (Kroll, 1998). É digno de nota que estas USs parecem ter surgido na família *Pasteurellaceae* e no gênero *Neisseria* independentemente (Findlay, 2009).

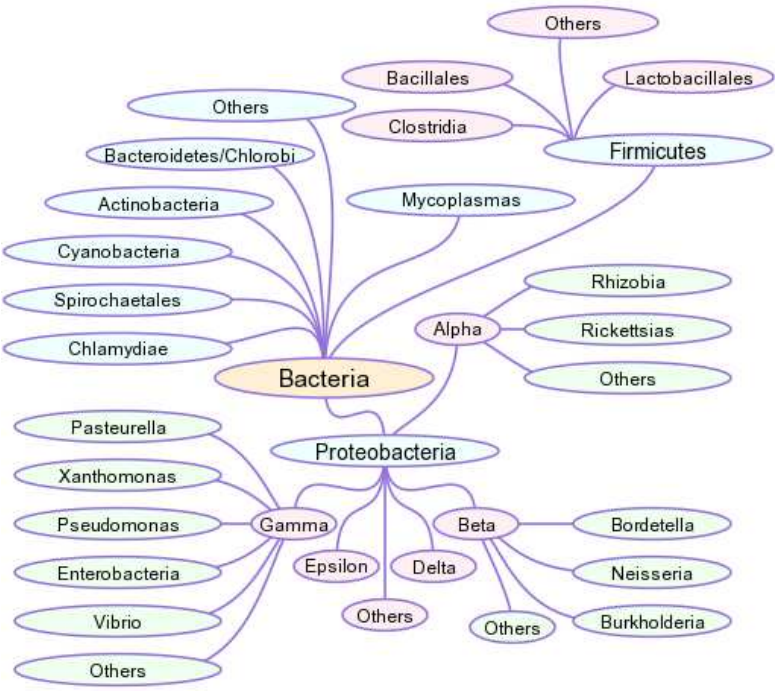


Figura 16: Classificação dos Filos pertencentes ao domínio *Bactéria.* Diagrama reproduzido de http://www.ncbi.nlm.nih.gov/sutils/genom_tree.cgi

Desta forma, uma alta frequência de troca de genes tem sido relatada, principalmente entre as espécies *H. infuenzae* e *N. meningitidis,* dada a grande quantidade de US específicas de *H. influenzae* presentes no genoma de *N. meningitidis* e vice-versa, sendo que a primeira via parece ser a via mais frequente (Kroll, 1998).

Vários processos de transferência horizontal têm sido identificados entre *N. meningitidis e H. influenzae*, que são espécies filogeneticamente distantes entre si, mas que compartilham o mesmo mecanismo de transformação como já exposto previamente. Em 1998, Kroll e colaboradores ao caracterizarem genes da cápsula em *N. meningitidis*, descobriram no genoma desta bactéria um segmento de DNA flanqueado por uma sequência US de *Haemophilus*e que possuía uma proporção anormalmente baixa de pares GC (cerca de 43%) para o genoma da espécie *N. meningitidis* (52% de pares GC) (Kroll,

1998). Ao se alinhar este segmento com outras sequências depositadas em bancos de dados, observou-se uma alta similaridade com o gene *sodC* da espécie *H. influenzae*. Árvores filogenéticas para este gene e outros genes destas espécies juntamente com a espécie Escherichia coli foram reconstruídas, evidenciando uma maior proximidade do gene entre as espécies *H. influenzae* e *N. meningitidis*. Desta forma, estes autores descreveram a transferência lateral do gene *sodC* de uma espécie do gênero *Haemophilus* para a *N. meningititis* (vale ressaltar que estes autores não afirmaram ser o *H. influenzae* a espécie doadora. Eles apenas utilizaram-na como base de comparação) (Kroll, 1998).

A importância da enzima SodC reside no fato da sua habilidade em promover o escape do sistema imune do hospedeiro, pela dismutação de ânions superóxido gerados externamente e que possuem alta afinidade por macromoléculas, especialmente DNA. Portanto, observa-se a troca de um fator de virulência entre gêneros bacterianos distintos e distantes filogeneticamente, o qual confere à bactéria receptora uma capacidade inteiramente nova: resistir à fagocitose, perpetuando o seu ciclo de vida no hospedeiro. Por conferir uma vantagem adaptativa ao meio, espera-se que este gene esteja sob uma forte pressão seletiva, principalmente do tipo adaptadora (Freeman, 2007).

Sabe-se que moléculas que constituem antígenos virais ou bacterianos, por exemplo, e, principlamente, aquelas que estão situadas na superfície do micro-organismo "aceitam" uma maior variabilidade na sua estrutura, o que permite um escape mais eficiente do sistema imune. (Aguileta, 2009). No entanto, o que se observa é que, apesar de amplamente distribuído entre as linhagens de *N. meningitidis* (presente em 100% das linhagens testadas) (Kroll, 1998), o gene está praticamente ausente entre as linhagens de *H. influenzae,* ocorrendo uma fraca prevalência de 17% entre os indivíduos não capsulados. Já entre os

capsulados, além deste gene ser ainda mais escasso, algumas linhagens possuem uma mutação na região do gene que codifica o sítio ativo da enzima, expressando uma enzima inativa, ao passo que outros chegam a não possuir o gene em seu genoma (Kroll, 1991).

Desta forma, o comportamento da SodC em linhagens capsuladas de *H. influenzae,* que são as mais invasivas, não sustenta a hipótese de que a SodC é necessária para a invasividade e virulência do organismo (Kroll, 1991). Em contrapartida, a constatação de que a SodC está amplamente presente em outros organismos comensais da orofaringe humana, sugere uma outra função, que não a de contribuir para a virulência e invasão do hospedeiro. Foi demonstrado que baixas concentrações de peróxido de hidrogênio levam a um rápido declínio da atividade ciliar do epitélio respiratório de ratos *in vitro* (Burman, 1986). Esta observação foi posteriormente extendida ao epitélio respiratório humano (Kroll, 1991). Especula-se, portanto, que a conversão *in loco* de ânions superóxido, provenientes de macrófagos residentes na orofaringe, para peróxido de hidrogênio pela SodC dificultaria o clearance bacteriano, prolongando a permanência da bactéria no hospedeiro. Desta forma, a SodC funcionaria muito mais como um fator de colonização do que de virulência (Kroll, 1991).

Estados como a bronquite crônica tem como característica um processo inflamatório leve, porém crônico, da mucosa respiratória produzido por uma superinfecção desta mucosa por micro-organismos comensais, como *H. influenzae* não capsulado e *H. parainfluenzae*. Nestes casos, a produção de ânions superóxido por células inflamatórias estará aumentada, e a sua rápida conversão em peróxido de hidrogênio pela SodC seria crucial para o escape bacteriano do clearance mucociliar (Kroll, 1991).

Além disso, constatou-se que bactérias altamente virulentas, como a *Neisseria gonorrhoeae*, comumente conhecida como gonococo, não possuem a enzima SodC, e a ausência desta enzima não afeta a virulência deste patógeno, mesmo num microambiente rico em reativos intermediários do oxigênio (Norrod & Morse, 1979). Olhando a situação sob este prisma, a SodC parece acrescentar uma vantagem seletiva à bactéria, mas não parece ser necessária para a sua sobrevivência, nem para a sua virulência. Uma outra hipótese é a de que esta enzime exiba um fenótipo extremamente plástico, funcionando como um fator de virulência em algumas espécies, como um fator de colonização, em outras, ou, ainda, sendo completamente desnecessária em outras espécies.

2. Justificativa

Desta forma, como a dinâmica do gene *sodC* é bastante distinta entre as espécies *H. influenzae* e *N. meningitidis,*as quais possivelmente protagonizaram o processo de transferência lateral deste gene (Kroll, 1998), especula-se que forças evolutivas distintas estejam moldando o processo de diversificação molecular deste gene nestas duas espécies bacterianas, uma possivelmente levando à fixação do gene na espécie *N. meningitidis,* e a outra conduzindo à perda do mesmo na espécie *H. influenzae.*

Como previamente exposto (Kroll, 1991; McCrea, 2008), existem espécies do gênero *Haemophilus,* como *Haemophilus haemolyticus* e *Haemophilus parainfluenzae,* membros comensais da orofaringe humana, que são muito próximas, filogeneticamente, do *H. influenzae,* sendo freqüente a troca de genes entre as mesmas (McCrea, 2008). Nestas espécies, o gene tem um comporatmento semelhante ao da espécie *N. meningitidis:* é amplamente distribuído entre as mais diversas linhagens e codifica um produto final ativo. Portanto, é razoável supor que a espécie *N. meningitidis* possa ter adquirido o gene *sodC* de uma outra espécie do gênero *Haemophilus,* e que nestas espécies o gene esteja sob um regime de forças seletivas semelhantes.

Faz-se, portanto, necessária a confecção e análise da topologia de uma árvore filogenética para o gene *sodC*, que agrupe outras espécies do gênero *Haemophilus* juntamente com a espécie *N. meningitidis.* Além disso, a realização de testes de Seleção Natural Positiva, quais forças estão moldando o processo de diversificação molecular nestas espécies, fornecendo indícios para a compreensão do desaparecimento do gene na espécie *H. influenzae.*

3. Objetivos

3.1 Objetivo Geral

Utilizar métodos de reconstrução filogenética, testes de seleção positiva e modelagem protéica para compreender o processo de diversificação molecular bem como os diferentes padrões de evolução do gene *sodC* nas bactérias naturalmente transformáveis *Haemophilus influenzae* e *Neisseria meningitidis.*

3.2 Objetivos Específicos

3.2.1 Alinhar sequências de nucleotídeos e estimar uma filogenia que permita apontar qual espécie do gênero *Haemophilus* compartilhou o gene *sodC* com a espécie *N. meningitidis;*

3.2.2 Aplicar um Teste de Seleção Positiva nos grupos de bactérias envolvidos no evento de transferência lateral no intuito de identificar os tipos de forças evolutivas subjacentes aos padrões de evolução e diversificação molecular do fator de virulência SodC;

3.2.3 Avaliar como estas forçaspodem estar moldando o comportamento do gene nas duas espécies bacterianas.

3.2.4 Modelar computacionalmente as possíveis novas formas da enzima SodC e avaliar as suas repercussões bioquímicas e funcionais.

2.2.5 Identificar as possíveis substituições de aminoácidos, suas implicações estruturais e funcionais, bem como sua repercussão na filogenia destes grupos, apontando quais os resíduos de aminoácidos cruciais para o evento de diversificação molecular do gene *sodC*.

4. Material e Métodos

4.1 Material

4.1.1 Meios de Cultura

4.1.1.1 Infusão de cérebro-coração (BHI) e BHI Ágar Suplementada

Para o crescimento de *H. influenzae*foi utilizada a forma comercializada na concentração de 37 g/L pela Acumedia(Lansing, Michigan, USA).O material, antes de ser usado,foi esterilizado em autoclave (esterilização por calor úmido) a 120 °C por 20minutos. Para o preparo do BHI Ágar, adicionou-se 1,5 % de ágar ao meio de cultura líquido. Além disso, ambos os meios líquido e sólido foram suplementados com NAD (2µg/ml) e Hemina (10µg/ml)

4.1.1.2 Ágar Chocolate

Para o crescimento de *N. meningitidis* foi utilizado o meio BHI Ágar, sendo adicionado5 % de Sangue Desfibrilado de Cavalo após o processo de autoclavagem, fazendo-se a ressalva de que esperava-se a temperatura baixar até atingir cerca de 80 °C, para evitar a desnaturação das proteínas do sangue.

4.2.1 Linhagens Bacterianas e Condições de Crescimento

Todas as linhagens de *Neisseria meningitidis*e *Haemophilus influenzae* utilizadas neste trabalho, bem como suas origens estão descritas nas tabelas 3 e 4. As linhagens foram cultivadas utilizando-se a técnica de esgotamento em Ágar Chocolate ou BHI Suplementadoà 37 °C por um período de 18-24 horas. Para a manutenção das amostras, as mesmas foram suspendidas em meio BHI ao qual foi adicionado glicerol a 30 %, e acondicionadas em biofreezer a -85 °C.

Tabela 3. Linhagens de *N. meningtidis* utilizadas neste trabalho

Linhagens N. meningitidis	Características	Origem
E.coli **DH5∝**	Nal^R utilizada como receptora de plasmídios. F-, *end*A1, *hsd*R17 (rk-, mk-), *sup*E44, *thi*-1, *gir* A96, *rel*A1	Hanahan et al.,1983
***N. meningitidis* P2143**	C:NT:P1.18-1,3 Cluster A4, ST8, liquor, Brasil - BA	INCQS - FIOCRUZ
***N. meningitidis* P2181**	C:NT:P1.5-1,10 Cluster A4, ST8, liquor, Brasil – BA	INCQS - FIOCRUZ
***N. meningitidis* P2200**	C:NT:P1.5-2,10 Cluster A4, ST153, liquor, Brasil – BA	INCQS - FIOCRUZ
***N. meningitidis* P2354**	B:NT:P1.7-2,3 Cluster A4, ST8, Líquor, Brasil – SC	INCQS - FIOCRUZ
***N. meningitidis* P2498**	C:NT:P1.5-2,10 Cluster A4, ST153, Líquor, Brasil – RJ	INCQS - FIOCRUZ
***N. meningitidis* B**	Sorogrupo B, ATCC 23247	INCQS - FIOCRUZ
***N. meningitidis* C2135**	Sorogrupo C, BIOMERIEUX C2135	INCQS - FIOCRUZ
***N. meningitidis* P374**	Sorogupo Y, Isolado de LCR Campo Grande – RJ	INCQS - FIOCRUZ

***N. meningitidis* W135**	Sorogrupo W135, ATCC35559	INCQS – FIOCRUZ
***N. meningitidis* A**	Sorogrupo A, ATCC 13077, líquor, caso fatal de meningite	INCQS – FIOCRUZ
N. meningitidis IA785101	*N. meningitidis*isolada em Corupá-SC. Sorogrupo C.	IAL – SP
N. meningitidis IA787101	*N. meningitidis*isolada em Corupá-SC. Sorogrupo B.	IAL-SP
***N. meningitidis* B-epi**	*N. meningitidis*epidêmica brasileira isolada pelo Instituto Adolfo LUTZ	IAL – SP
***N. meningitidis* Y**	*N. meningitidis*isolada em surtos epidêmicos de meningite nos Estados Unidos	USA
***N. meningitidis* B4**	*N. meningitidis*B4:P1-16 padrão.	IAL – SP
***N. meningitidis* W135 ref**	*N. meningitidis*de referência do sorogrupo W135.	IAL – SP
***N. meningitidis* W135 Marseille**	*N. meningitidis* do sorogrupo W135 isolada na cidade de Marseille-France proveniente do surto W135 ST11 (ET37) Hajj.	IP

IAL – Instituto Adolfo Lutz SP

INCQS – FIOCRUZ – Instituto Nacional de Controle de Qualidade- Fundação Oswaldo Cruz.

LBMB – Laboratório de Biologia Molecular Bacteriana – IB – UNICAMP

IP – França – Centre National de Référence des Ménigocoques – Institut Pasteur De Paris – França

Tabela 4. Linhagens de *H. Influenzae* utilizadas neste trabalho.

Linhagens H. influenzae	Características	Origem
RdKW20	*Haemophilus influenzae* cepa padrão sequenciada completamente, do sorotipo d	INCQS – FIOCRUZ
Hib-βlac	*Haemophilus influenzae* sorotipo b ATCC *β Lactamase* positivo	INCQS – FIOCRUZ
Hi01	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	

Hi02	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi03	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi05	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi07	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
		LBMB
Hi07	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi08	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi09	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi10	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi11	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi12	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
	Haemophilus influenzae NTHi isolado de um caso de fibrose cística	LBMB
Hi13	*Haemophilus influenzae*NTHiisolado de um caso de fibrose cística	LBMB
Hi19	*Haemophilus influenzae NTHi* isolado de um caso de fibrose cística	LBMB
Hi20	*Haemophilus influenzae*NTHiisolado de um caso de bronquite	LBMB
Hi21	*Haemophilus influenzae*NTHiisolado de um caso de amigdalite	LBMB
Hi22	*Haemophilus influenzae*NTHiisolado de um caso de amigdalite	LBMB
Hi23	*Haemophilus influenzae*NTHiisolado de secreção brônquio-alveolar	LBMB
Hi24	*Haemophilus influenzae*NTHiisolado de secreção brônquio-alveolar	LBMB
Hi26	*Haemophilus influenzae*NTHiisolado de secreção ocular	LBMB
Hi27	*Haemophilus influenzae*NTHiisolado de secreção ocular	LBMB
Hi28	*Haemophilus influenzae*NTHiisolado de secreção ocular	LBMB
Hi29	*Haemophilus influenzae*NTHiisolado de um caso de sinusite	LBMB
Hi30	*Haemophilus influenzae*NTHiisolado de líquido sinovial	LBMB

Hi31	*Haemophilus influenzae*NTHiisolado de um caso de sinusite	LBMB
Hi32	*Haemophilus influenzae*NTHiisolado de secreção brônquica	LBMB
Hi33	*Haemophilus influenzae*NTHiisolado de secreção brônquica	LBMB
Hi34	*Haemophilus influenzae*NTHiisolado de secreção brônquica	LBMB
Hi35	*Haemophilus influenzae*NTHiisolado de escarro	LBMB
Hi36	*Haemophilus influenzae*NTHiisolado de um caso de pneumonia	LBMB
Hi37	*Haemophilus influenzae*NTHiisolado de um caso de pneumonia	LBMB
Hi38	*Haemophilus influenzae* sorotipo aisolado de um caso de pneumonia	LBMB
Hi39	*Haemophilus influenzae*NTHiisolado de um caso de meningite	LBMB
Hi40	*Haemophilus influenzae*NTHiisolado de um caso de meningite	LBMB
Hi41	*Haemophilus influenzae*NTHiisolado de um caso de meningite/LCR	LBMB
Hi42	*Haemophilus influenzae* sorotipo b isolado de um caso de meningite/LCR	LBMB
Hi43	*Haemophilus influenzae* sorotipo b isolado de um caso de meningite/LCR	LBMB
Hi44	*Haemophilus influenzae*NTHiisolado de um caso de líquido pleural	LBMB
Hi45	*Haemophilus influenzae* sorotipo b isolado de uma hemocultura/meningite	LBMB
Hi46	*Haemophilus influenzae* sorotipo b isolado de uma hemocultura/meningite	LBMB
Hi47	*Haemophilus influenzae* NTHi isolado de uma hemocultura/meningite	LBMB
Hic	*Haemophilus influenzae* sorotipo c ATCC 9007 NCTC 8469.	IAL – SP
Hie	*Haemophilus influenzae* sorotipo e NCTC 10479.	IAL-SP
Hif	*Haemophilus influenzae* sorotipo f NCTC 7918.	IAL – SP

IAL – Instituto Adolfo Lutz SP

INCQS – FIOCRUZ – Instituto Nacional de Controle de Qualidade- Fundação Oswaldo Cruz.

LBMB – Laboratório de Biologia Molecular Bacteriana – IB – UNICAMP.

4.2 Métodos

4.2.1 Extração de DNA

As linhagens de interesse foram cultivadas em meio sólido, a 37 C por um período de 18 a 24h. Para a extração de DNA cromossomal bacteriano, uma amostra bacteriana correspondente a uma DO em 600nm= 2.0 foi ressuspendida em tampão TE (Tris-EDTA). A técnica empregada foi a do tampão CTAB (Ausubel, 1988) (Tabela5). A integridade do DNA obtido foi avaliada através de eletroforese em gel de agarose a 0.7% (Sambrook, 1989).

Tabela 5: Composição do Tampão CTAB

Composição do Tampão CTAB
TrisHCl a 100mM, pH= 8
EDTA a 20mM
NaCl a 1.4M
β-mercaptoetanol a 0.2%
Proteinase K a 10mg/ml

4.2.2 Amplificação do Gene de Interesse por PCR

Para a amplificação do gene *sodC* nas espécies *N. meningitidis* e *H. Influenzae*, foram confeccionados 3 pares de *primers*: 1 par para as linhagens de *N. meningitidis*, 1 par para as linhagens de *H. Influenzae* NTHi e 1 par para as linhagens capsuladas tipáveis de *H. Influenzae* (tabela 6). Os moldes utilizados, bem como os respectivos números de acesso

estão sumarizados na mesma tabela. O programa Gene Runner 3.01 (Spruyt, 1994) foi utilizado para desenho dos primers e cálculo da temperatura de anelamento. Em seguida, a opção BLAST 2seq. da ferramenta BLAST disponibilizada no site www.ncbi.nlm.nih.gov foi utilizada para avaliação da especificidade de cada par em relação ao seu respectivo alvo ou molde, bem como com relação a outras sequências mais similares.

Tabela 6. Primers utilizados na reação de PCR e na reação de sequenciamento. AC = *accession number*

ESPÉCIE	*MOLDE*	*AC*	*TAMANHO DO PRODUTO*	*SEQUÊNCIA*
N. meningitidis	**053442**	**CP000381.1**	**555pb**	**F 5-TGAAAACCTTATTAGCACTAGC-3** **R 5-TATTTAATCACGCCACATGC-3**
H. influenzae **Capsulado**	**Hib**	**M84012.1**	**563pb**	**F 5-ATGATGAAAATGGAAAACTCTCTTAG-3** **R 5-TTATTTAATCACGCCACATGC-3**
H. influenzae **NTHi**	**P18-25**	**GU269224.1**	**516pb**	**F 5-GTTAGTGCGGTATGTTCAGTTG-3** **R 5-CGCCAAGTGGAGCAGGAT-3**

Na reação de amplificação foram utilizados 200ng de DNA genômico, para um volume final de 50µL. As condições padrão da reação estão sumarizadas na tabela 7. As temperaturas de anelamento (TM) estão descritas na tabela 8.

A amplificação do produto de PCR das linhagens de *N. meningitidis* seguiu a reação padrão. Como poucos produtos foram obtidos à partir das linhagens de *H. Influenzae,* variações de 3o C na temperatura de anelamento foram feitas para mais e para menos, sendo retestadas todas as amostras não amplificadas. Nenhum produto adicional foi obtido.

Tabela 7: Condições padrão da reação de PCR

Condições padrão da reação de PCR		*Ciclos N. meningitidis/ H. influenzae*	
Tris-HCl, pH= 8.4	**10mM**	**95o C**	**5min**
MgCl2	**1.5mM**	**95o C**	**1min**
Primers	**50pmol (cada)**	**TM**	**1min**
dNTPs	**0.2mM (cada)**	**72o C**	**1min**
***Taq* DNA Polimerase**	**2.5U**	**72o C**	**10min**
		Ciclos totais = 30	

Tabela 8: Temperatura de anelamento para cada espécie bacteriana.

Espécie	**TM**
N. meningitidis	**60o C**
H.influenzae NTHI	**55o C**
***H. influenzae* capsulado tipável**	**58o C**

4.2.3 Reação de Sequenciamento do Produto de PCR

Para a reação de sequenciamento do respectivo produto de PCR foram utilizados os mesmos pares de primers.

Tabela 9: Linhagens bacterianas sequenciadas à partir do LABIOTEC.

N. meningitidis	*Haemophilus influenzae*
A_ATCC	**Hi 12 (NTHi)**
B_ATCC	**Hi 19 (NTHi)**
B4	**Hi e (capsulado sorotipo e)**
B epidêmica	
IAL 2443	

P2354
P2143
P2181
P2200
P2498
C2135
Marseille
W135
Y USA
P374

Optou-se pelo BigDye Terminator v3.1 Cycle Sequencing Kit (Life Technologies) com corrida em sequenciador 3500 Genetic Analizer (Life Technologies), segundo recomendações do fabricante. As sequências obtidas foram comparadas com as sequências de referência (molde), depositadas no Genebank e descritas na Tabela 5, através da ferramenta Blast 2q, disponível no site do NCBI.

A edição e montagem das sequências foi deita utilizando-se os programas BioEdit 7.0.9.1 (Hall,1997) e Chromas*Pro* 1.5 (Technelsium Pty Ltd, Tewantin QLD 2003, Austrália). A seuência de ida foi alinhada ao reverso complementar da sequência de volta através do algorítmo de alinhamento global. Os pareamentos dúbios foram resolvidos analisando-se o cromatograma. O pico com melhor nota *Q* para uma dada posição foi considerado o melhor pareamento. Por terem baixa qualidade, pares do início e do final de cada sequência foram descartados (Ewing, 1998).

4.2.4 Obtenção de Sequências Homólogas à partir do GeneBank

A ferramenta BLASTN *(Basic Local Alignment Search Tool)*(Altschul, 1990, Altschul, 1997) implementada no site do NCBI (*National Center for Biotecnology Information,* EUA – www.ncbi.nlm.nih.gov) foi utilizada para se obter sequências no

formato FASTA à partir do Banco de Dados nr/nt (*Nucleotide Collection*) usando-se as configurações padrão do MEGABLAST. Os números de acesso das sequências pergunta (*query sequences*) para as espécies *H. influenzae* e *N. meningitidis* estão dispostos na tabela 10 Para a obtenção de sequências de outras espécies dentro da mesma família, foram utilizados os mesmos números de acesso, porém usando-se o DISCONTIGUOUS MEGABLAST sob as mesmas configurações padrão. Na mesma tabela estão dispostos os números de acesso das sequências utilizadas neste trabalho.

Tabela 10: Números de acesso (AC) : utilizados neste trabalho.

ESPÉCIES	*NÚMEROS DE ACESSO (AC)*
N. meningitidis	**GQ365733.1 a GQ365746.1; CP002419.1 a CP002424, CP000381.1 e CP001561.1; AM889136.1, FR774048.1, FM999788, AL157959.1, FR845709.1, AM421808, AE002098.2**
H. influenzae NTHI	**GU269206.1 a GU269227.1**
H.influenzae	**M84012.1 (HiB)**
	AF549211.1 (HiF)
H. parainfluenzae	**M84013.1 e FQ312002.1**
H. haemolyticus	**GU269205.1 e GU269211.1**
H. ducreyi	**. AE017143.1, U47664.1 e X98737.1**
Actinobacilus. pleuropneumoniae	**X83123.1**
Yersinia pestis	**CP000901.1**

4.2.5 Alinhamento das Sequências de Nucleotídeos

Um dos propósitos da análise filogenética é estabeler as relações evolutivas entre grupos de genes através da inferência da ancestralidade comum entre os mesmos, e gerar suposições acerca de como forças evolutivas moldaram suas sequências, gerando sua diversificação molecular. Para isto, é essencial que sítios homólogos sejam comparados uns com os outros. Assim, ao se fazer um alinhamento de sequências de nucleotídeos, as sequências homólogas sob investigação são alinhadas de forma que os sítios homólogoso fiquem dispostos em colunas. Para uma dada coluna, o sítio referente àquela posição pode permanecer igual ao seu estado ancestral ou mudar de estado com uma dada probabilidade (Lemey, 2009).

Diante de *indels* (o nome indel é aplicado para se referir a inserções e/ou deleções, quando não se sabe a origem do evento, se inserção em uma sequência ou deleção na outra (Lemey,2009)), gaps (ou lacunas) são introduzidos, de forma que as sequências permaneçam do mesmo tamanho, garantindo, portanto, que sítios homólogos permaneçam alinhados corretamente. A lacuna corresponde a uma penalidade. Toda vez que o programa introduz uma lacuna, ele está penalizando aquele alinhamento. O número de lacunas entre duas sequências é um dos parâmetros utilizados pelo programa para calcular o grau de similaridade entre as sequências, o que permite ao observador supor se existe ou não uma hipótese de homologia entre as mesmas. Portanto, quanto maior o número de lacunas, menor a similaridade entre as sequências, e, portanto, mais distantes elas são umas das outras e vice versa.

É muito comum que fragmentos de tamanhos diferentes do mesmo gene sejam sequenciados.Quando sequências de tamanhos diferentes são alinhadas, o programa irá alinhar os sítios homólogos e, nas extremidades que não foram sequenciadas, ele irá

introduzir lacunas, portanto estará penalizando aquele alinhamento, de modo que uma hipótese errônea de homologia entre duas sequências pode ser gerada, não porque elas não sejam similares entre si, mas por que dados (informação) estão faltando. Uma forma de corrigir este erro é retirar manualmente do alinhamento as regiões de tamanhos diferentes, deixando, todas as sequências do mesmo tamanho, para que a mesma quantidade de informação seja analisada em todas as sequências. Esta alternativa, que é a mais comumente utilizada, pode diminuir o poder do alinhamento (ao se retirar sítios do alinhamento estamos retirando parâmetros que suportam ou rejeitam uma hipótese de homologia). Desta forma, deve-se sempre tentar alinhar o maior número possível de sítios entre duas sequências. Além disso, ao se fazer um alinhamento parcial, pode-se incorrer no erro de perder regiões informativas, por exemplo: consideremos que as extremidades não sequenciadas correspondam a porções onde as sequências apresentem alta variabilidade.Portanto duas sequências podem exibir um alto grau de similaridade porque as regiões variáveis do gene não foram consideradas. O contrário também é possível. Desta forma, quanto menor o número de sítios considerados, maior a chance de se gerar um pareamento enviesado, o que poderá criar uma falsa suposição de homologia.

Alinhamentos são construídos utilizando-se pacotes computacionais, os quais implementam algorítmos específicos. As sequências de nucleotídeos foram alinhadas utilizando-se os parâmetros padrão do programa CLUSTALW (Thompson, 1997) implementado no pacote computacional MEGA 5.05 (Kumar, 2000). O CLUSTALW implementa um algorítimo bastante popular conhecido como Alinhamento Progressivo. Este inicia sua análise comparando a similaridade entre todos os pares de sequencias, alinhando primeiramente o par com a maior nota de similaridade. Este par é então tratado como uma única sequência, e a sequência seguinte mais similar a esta é adicionada.

Portanto, as outras sequências são adicionadas progressivamente, em ordem de similaridade. (Figura 17).

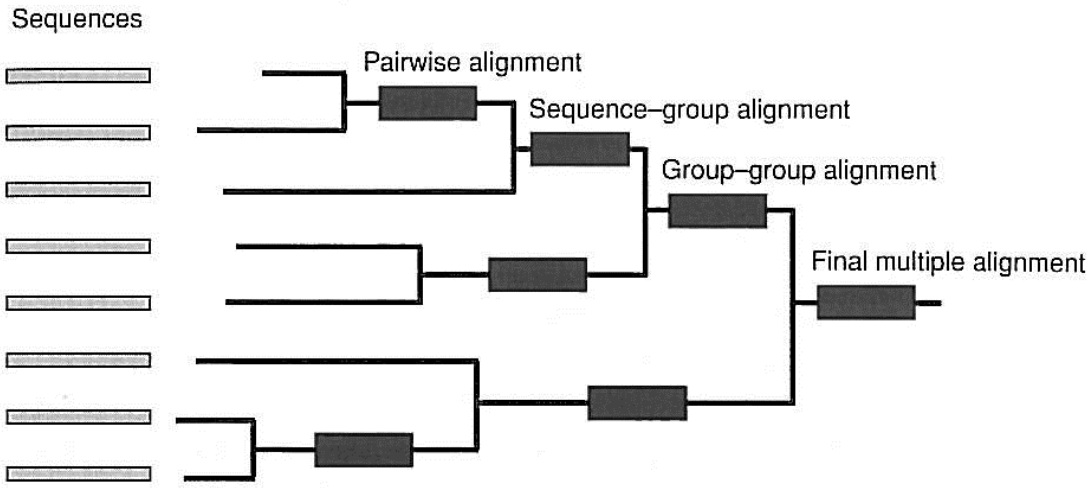


Figura 17: Etapas de um alinhamento progressivo (Figura reproduzida de Lemey, 2009).

O programa BioEdit 7.0.9.1 (Hall,1997) foi usado para fazer a conversão do formato das sequências de FASTA para PHYLILP, permitindo o seu uso no pacote PAML 3.14 *(Phylogenetic Analysis by Maximum Likelihood)* (Yang, 2004).

4.2.6 Estimação e Reconstrução de Árvores Filogenéticas para o gene *sodC*

A reconstrução de uma árvore filogenética baseada em dados de sequências moleculares é, simplificadamente, baseada em três etapas: 1) alinhamento de múltiplas sequências, 2) escolha do modelo de substituição de nucleotídeos (ou modelo evolutivo), e 3) escolha do modelo estatístico de estimação da topologia de árvore (Lemey, 2009).

Um modelo de substituição de nucleotídeos calcula a probabilidade de um dado nucleotídeo numa dada posição permenecer no mesmo estado ou mudar de estado.

Tal modelo deve ser empregado para corrigir o fenomêno de Saturação: quando duas sequências já estão divergindo há muito tempo, observa-se que, em certas posições, duas ou mais substituições ocorreram mudando o estado de carate do sítio ou não. O investigador, observando o fenômeno à partir do momento presente, no entanto, somente é capaz de observar um único evento, ou nenhum (se a substituição retomou o estado ancestral do sítio em questão). Portanto, pode-se incorrer no erro de se atribuir um valor de similaridade para um dado par se sequências maior do que o seu valor real.(Lemey, 2009)

Os modelos estatísticos utilizados para a reconstrução de árvores filogenéticas à partir de dados moleculares podem ser classificados de acordo com o tipo de dado utilizado, e de acordo com o algorítimo implementado (Tabela 11). De acordo com o tipo de dado, temos: métodos baseados em estados de caracteres discretos, ou métodos baseados em matrizes de distância. De acordo com o algorítimo implementado pelo método, temos aqueles baseados em agrupamento, cujo resultado final é uma única topologia de árvore, e aqueles baseados num critério ótimo, o qual é usado para buscar a melhor árvore, num conjunto final de possíveis topologias de árvores (Lemey, 2009).

Tabela 11: Métodos para se estimar árvores filogenéticas e suas estratégias (Tabela reproduzida de Lemey, 2009).

	CRITÉRIO ÓTIMO	*AGRUPAMENTO*
ESTADO DE CARATER	Máxima Parsimônia Verossimilhança Máxima Inferência Bayesiana	
MATRIZ DE DISTÂNCIA	Fitch-Margoliash	UPGMA Junção de Vizinhos

As filogenias foram estimadas à partir de sequências de DNA utilizando-se o método de Junção de Vizinhos (NJ – *Neighbor Joining*). Este método é baseado na construção de uma matriz de distância de dissimilaridades par a par, e emprega um algorítimo de agrupamento, resultando numa única topologia final possível de árvore.

Os métodos baseados em matrizes de distância agrupam progressivamente os pares de sequência com as maiores notas de similaridade e, posteriormente, inferem as relações filogenéticas entre os taxons à partir desta matriz. Os algorítimos de agrupamento vão gradualmente os taxons numa única árvore final, evitando, desta forma, o problema de ter que se avaliar um grande número de possíveis topologias finais de árvore. Existem diversos algorítimos de agrupamento que diferem no método empregado para determinar a relação entre os taxons, na maneira de agrupar os taxons, e no cálculo do tamanho dos ramos (Lemey, 2009).

O método da Junção de Vizinhos, em particular, inicia a análise assumindo uma árvore com formato de estrela, na qual não existem ramos internos. O primeiro ramo interno, compostos por um par de sequências, é introduzido a esta árvore e a nota da árvore é então calculada. Depois todos os possíveis pares de sequências são adicionados sequencialmente, de forma a gerar a árvore com a menor nota. O método não almeija agrupar as sequências mais proximamente relacionadas, mas sim, minimizar o comprimento de todos os ramos internos, gerando, assim, a árvore com a menor nota (Lemey, 2009).

O modelo de substituição de nucleotídeos implementado foi o *Maximum Composite Likelihood,* levando-se em consideração os seguintes parâmetros: transições e transversões são estados distintos, as taxas de mutação são uniformes entre os sítios e padrões homogêneos entre linhagens.

O teste de confiabilidade amostral escolhido foi o de Reamostragem com Reposição (*Bootstrap*) com 1000 repetições. Este teste reconstrói a árvore muitas vezes re-amostrando os dados iniciais, de forma que o valor apresentado é a porcentagem de vezes que a árvore com a mesma topologia foi recuperada durante o processo de re-amostragem. A árvore consensus foi gerada utilizando-se um Treshold de 50%. Tal procedimente é implementado pelo programa MEGA 5.05. (Kumar, 2000).

4.2.7 Análise da Ocorrência de Seleção Positiva (Teste de Seleção)

Inicialmente foi feito o alinhamento das sequências de interesse, e uma árvore filogenética para estas sequências foi estimada através de um método estatístico de distância (*Neighbor Joining* – Junção de Vizinhos). O modelo de substituição de nucleotídeos foi o de Tamura-Kumar com taxas de mutação uniformes entre os sítios e padrões homogêneos entre linhagens. O teste de confiabilidade estatística utilizado foi o de Reamostragem com Reposição (*Bootstrap*) com 1000 repetições e o treshold foi de 50%. O alinhamento e a árvore filogenética serviram, respectivamente, como arquivos de entrada para o programa.

Para testar a presença de seleção positiva, foi utilizado o programa CODEML implementado pelo pacote computacional PAML. Este programa utiliza métodos estatísticos para detectar seleção molecular adaptativa comparando as taxas de substituições não sinônimas (*dN* - mutação de um nucleotídeo no codon que gera troca do aminoácido correspondente) e de substituições sinônimas (*dS* - mutação de um nucleotídeo no codon que não gera troca do aminoácido correspondente, portanto, é dita silenciosa) ao longo de uma sequência de DNA que codifica uma proteína, portanto o programa baseia-se no cálculo do valor de ω, que é a razão entre estas taxas: $\omega = dN/dS$. Quando ω>1, seleção

positiva está operando na sequência. Quando ω<1, seleção negativa está atuando na sequência e, finalmente, se ω =1, substituições sinônimas e não sinônimas ocorrem com a mesma chance ao longo de uma dada sequência, portanto não existe um regime de força determinística operando sob a mesma. Diz-se que a sequência encontra-se sob evolução neutra.

Para o cálculo do valor da razão ω, primeiramente o programa implementa um modelo de substituição de codons, descrito por Goldman e Yang em 1994 (Yang, 1994). Este modelo especifica que a taxa de modificação do codon no estado *i* para o estado *j* se dá da seguinte maneira:

$$q_{ij} = \begin{cases} 0, & \text{Se dois códons diferem em mais de uma posição} \\ \pi_j, & \text{Para transversões sinônimas} \\ \kappa\pi_j, & \text{Para transições sinônimas} \\ \omega\pi_j, & \text{Para transversões não sinônimas} \\ \omega\kappa\pi_j, & \text{Para transições não sinônimas} \end{cases}$$

Figura 18: Possibilidades de mudança de estado do codon (Figura modificada de Yang, 1997)

O codon no estado *j* pode ser considerado um parâmetro livre ou pode ser calculado à partir das frequências de nucleotídeos nas três posições do codon. O método estatístico de Processos de Markov é implementado para modelar a substituição de codons. Através deste modelo que cada sítio (neste caso o codon) pode variar livremente e independentemente dos outros. Em seguida, é realizado o cálculo de ω para cada sítio comparando-se duas sequências. O modelo de Processos de Markov permite que a razão ω varie entre os ramos

de uma filogenia ou entre os sítios de uma sequência. Depois disto, o Teste da Razão Verossimilhanças é aplicado para se comparar a performance entre pares de modelos: um modelo do par testa para a neutralidade (hipótese nula) e o outro modelo do par testa para seleção positiva (hipótese alternativa). O Teste da Razão Verossimilhanças (LRT – *Likelihood Ratio Test*) permite identificar algumas linhagens sob seleção positiva em uma árvore filogenética, ou pode inferir quais resíduos críticos de aminoácidos numa proteína estão sob seleção diversificadora.

Os modelos de ramos (*branch models*) permitem que o valor da razão ω varie entre os ramos de uma dada árvore filogenética e são úteis para se detectar seleção somente entre algumas linhagens de uma filogenia (Yang, 1998; Nielsen, 1998). Tais modelos são conhecidos como MA nulo (neutro) e MA alternativo (seleção positiva).

Os modelos de sítios permitem que ω varie entre os codons de uma sequência de DNA (Nielsen, 1998; Yang, 2000). Para cada modelo, com exceção do modelo M0, o valor de ω assume o carater de uma distribuição de probablidades de valores distintos de ω de acordo com o tipo e a quantidade de sítios que são fixados como parâmetro, por exemplo no modelo M2a existem três categorias de sítios: p0 que, assume-se, estão sob seleção negativa, portanto para os sítios p0, $\omega<1$; p1(p1= 1 – p0) que são neutros. Nestes sítios $\omega=1$; e p2 (p2 = 1 – p0 – p1) que estão sob seleção positiva. Neles, $\omega>1$.

Cada modelo tem uma nota ou *score*. Essas notas são comparadas par a par através do Teste da Razão de Verossimilhanças (um modelo do par assume a hipótese nula e o outro assume a hipótese alternativa – seleção positiva atuando), e a significância estatística desta comparação é dada pelo teste do Qui-quadrado. Os graus de liberdade correspondem à diferença entre o número de parâmetros livres estimados entre cada modelo. Dois pares são paticularmente úteis: M1a (modelo aproximadamente neutro - que pressupõe uma

proporção p0 de sítios conservados com ω0<0, e p1= 1-p0 de sítios neutros com ω1=1, de forma que todas as mutações são consideradas deletérias ou neutras) X M2a (modelo de seleção positiva - que apresenta uma classe adicional de sítios, sendo que p2= 1-p0-p1 e ω2 é estimado à partir dos dados), e M7 (Distribuição Beta - o qual é um teste nulo para seleção positiva, correspondendo a uma distribuição beta que pode assumir valores no intervalo 0<ω<1) X M8 (Beta&ω - com quatro parâmetros livres (p, q, p0 e ω), que adiciona uma classe discreta ω ao modelo beta (M7), identificando os sítios submetidos a seleção positiva (ω>1)).

Tabela 12: Modelos de sítio e seus respectivos parâmetros. #p corresponde ao número de parâmetros livres na distribuição ω. Os parâmetros entre parênteses não são considerados livres (Tabela reproduzida de Yang, 1997). Model = modelo ; sites = sítios, Parâmetros = Parâmetros.

Model	NSsites	#p	Parameters
M0	0	1	ω
M1a	1	2	p_0 $(p_1 = 1 - p_0)$, $\omega_0 < 1$, $\omega_1 = 1$
M2a	2	4	p_0, p_1 $(p_2 = 1 - p_0 - p_1)$, $\omega_0 < 1$, $\omega_1 = 1$, $\omega_2 > 1$
M3	3	5	p_0, p_1 $(p_2 = 1 - p_0 - p_1)$ $\omega_0, \omega_1, \omega_2$
M7	7	2	p, q
M8	8	4	p_0 $(p_1 = 1 - p_0)$, $p, q, \omega_s > 1$

O modelo M0 (um único valor de ω) é um modelo nulo e calcula um único valor de ω para toda a filogenia. Nele ω<1. Este pode ser comparado com o modelo alternativo M3 (Modelo Discreto - no qual as probabilidades p0, p1 e p2 de cada sítio de cada sítio que está submetido a seleção (purificadora, neutra ou positiva, respectivamente) e seus

correspondentes valores de ω (ω0, ω1 e ω2) são inferidos à partir dos dados), o qual não é um modelo de seleção propriamente dito, porém avalia a diversidade de seleção ao longo de uma sequência. Este modelo assume três categorias de sítios distintos e diz se há mais de um tipo de seleção atuando numa sequência. O par M1a X M2a parece ter menor poder para detectar seleção positiva do que o par M7 X M8, segundo Nielsen e Yang, 1998. Resumindo os modelos de sítio são usados para testar diferentes distribuições de probabilidade para ω heterogêneo entre os códons possibilita a identificação daqueles códons que estão sendo selecionados positivamente, através da discretização das probabilidades em diferentes classes de códons (Yang, 1997; Yang, 2000; Yang, 2002).

Se o teste da razão de Máxima Verossimilhança para um par é significativo, ou seja, 3.84 para 5%, então o método de Bayes Empirical Bayes (BEB) é empregado para calcular a probabilidade *a posteriori* de cada classe de sítio, permitindo a identificação dos resíduos sob seleção positiva (Yang, 2005).

Foram testados 3 modelos de sítio (M0, e M1a X M2a) e os dois modelos de ramo (MAnulo X MAalternativo) para ramos especificamente selecionados na árvore.

4.2.8 Modelagem Proteica Computacional por Homologia

A modelagem molecular da estrutura da enzima SODC de *H. haemolyticus* foi feita utilizando-se o programa YASARA (Yer Another Scientific Artificial Reality Application) (Krieger, 2009).Esta espécie foi a escolhida por ser a provável doadora para as espécies *N. meningitidis* e *H. influenzae* e por codificar uma enzima sabidamente ativa (McCrea, 2010). Além disso, o gene é amplamente difundido entr as suas linhagens (Fung, 2006).

O programa recebe uma sequência de aminoácidos como entrada e retorna um modelo refinado de alta resolução, usando um protocolo aprovado pela CASP (Critical

Assessment of Structure Prediction) (Krieger, 2009). O molde utilizado foi a estrutura da [Cu,Zn] - Superóxido Dismutase de *Neisseria meningitidis,* 2AQN, segundo PDB (RCSB Protein Data Bank – www.rcsb.org/pdb).

A estrutura resultante foi submetidas a minimização de energia e simulação de dinâmica molecular usando o programa YASARA. Para isso, todos os átomos de hidrogênio e outros átomos ausentes do modelo foram criados usando parâmetros do campo de força YAMBER3. Uma caixa de simulação foi definida em 15 Å em torno de todos os átomos de cada complexo. Protonação foi realizada com base no pH 7. Neutralização da célula foi alcançada preenchendo a caixa com moléculas de água e íons Na/Cl. Uma curta simulação de dinâmica molecular foi realizada para ajustar o solvente, excluindo moléculas de água até que a densidade de água de 0,997 g/ml fosse atingida. Uma minimização de energia usando o algoritmo 'steepest descent' foi realizada até que a velocidade máxima dos átomos caísse abaixo de 2200 m/s.

5. Resultados e Discussão

5.1 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC* de *N. meningitidis*.

Genes são ditos homólogos quando eles compartilham ou descendem de um estado ancestral comum, a qual sofreu um processo de diversificação ao longo do tempo, dando origem à variabilidade observada entre as sequências descendentes ou contemporâneas. Similaridade é a forma de se quantificar a homologia entre genes, contando-se o número de nucleotídeos idênticos relativo ao tamanho da sequencia. Portanto, quanto maior o número de bases idênticas entre sequências, mais similares elas são entre si. Sendo assim, a hipótese de homologia estará correta (Lemey, 2009)

Apesar de se assumir que duas sequências evoluíram de um estado ancestral comum, ao longo do tempo, a similaridade entre elas pode ter sido "apagada" de tal forma que elas já não carregam mais informação suficiente acerca do seu estado ancestral. Portanto, o termo homologia é usado apenas quando o ancestral comum é recente o bastante na escala de tempo para que as sequências descendentes retenham similaridade o suficiente, de forma que possam usadas para fins de análise filogenética (Lemey, 2009).

Partindo-se do pressuposto de que cada sequência "conta uma história", alinhar múltiplas sequências é confeccionar uma hipótese de homologia sobre múltiplos resíduos de aminoácidos numa proteína ou sobre sítios numa sequência de nucleotídeos. Portanto, assume-se que cada resíduo ou sítio alinhado divergiu de um estado ancestral comum. Desta maneira, é de suma importância que sítios homólogos sejam comparados uns com os outros, certificando-se o investigador de que não está comparando sítios parálogos. Assim,

as sequências sob investigação devem ser alinhadas de tal maneira que os sítios homólogos formem colunas ao longo do alinhamento (Lemey, 2009)

Existem inúmeras razões para se fazer um alinhamento de múltiplas sequências. Uma delas é demonstrar a similaridade numa família de sequências. A mais comum, no entanto, é que o alinhamento de múltiplas sequências fornece o substrato para a reconstrução de uma filogenia. Ao se reconstruir uma filogenia de sequencias codificantes de nucleotídeos, o investigador tem por finalidade avaliar as relações evolutivas entre grupos específicos de genes (quais sequencias são mais proximamente relacionadas entre si) e/ou inferir acerca de hipóteses de ancestralidade comum entre esses grupos, além de testar hipóteses acerca dos processos que moldaram a sequência de tais genes ao longo da sua história evolutiva. Neste caso, diz-se que a árvore em questão é uma árvore de genes, pois ela "conta" a história de genes específicos e não de espécies. Para que uma árvore de genes reflita a história evolutiva de uma espécie, os genes em questão devem ser ortólogos, e não parálogos, além de serem conservados. Geralmente, uma árvore de espécies com base em dados moleculares é feita à partir de 5 a 7 genes *housekeeping* concatenados (McCrea, 2008; McCrea, 2010)

Assumindo-se que todas as sequências evoluíram de um estado ancestral comum, a diversidade observada entre as sequências contemporâneas é, desta forma, fruto de diferentes mecanismos moleculares responsáveis por introduzir variabilidade ao genoma (mutação de ponto, inserção, deleção, duplicação, recombinação, etc). (Lemey, 2009)

Durante a duplicação da informação genética, a DNA polimerase pode, ocasionalmente, incorporar uma base não complementar. Além disso, as bases do DNA podem ser quimicamente modificadas devido a fatores ambientais, tais como: radiação UV, substâncias químicas, etc. Quando estas modificações escapam aos mecaqnismos de reparo

celular, a sequência de DNA á alterada, resultando na chamada mutação de ponto. O código genético evoluiu de tal maneira que uma mutação de ponto na terceira posição do códon raramente resulta em modificação do aminoácido correspondente (apenas 30% de chance). Na segunda posição do códon, uma mutação de ponto sempre resulta em modificação do aminoácido, e na primeira posição, ocorre modifiocação em 96% das vezes. Mutações (ou substituições de bases) que não resultam em modificação do aminoácido à nível da proteína são ditas mutações silenciosas ou substituições sinônimas. Quando uma mutação resulta na incorporação de um aminoácido diferente, ela é dita substituição não sinônima (Lemey, 2009).

A comparação entre as sequencias de genes homólogos revelou que genes de espécies mais proximamente relacionadas geralmente diferem apenas por um númaro limitado de mutações de ponto. Além disso, estas são encontradas preferencialmente na terceira posição do códon, resultando numa taxa de evolução mais rápida nesta posição, comparada com as demais posições. A redundância do código genético faz com que genes possuam uma taxa de evolução maior do que as proteínas que eles codificam (Lemey, 2009).

Erros de replicação do DNA também podem resultar em deleções ou inserções de um ou mais nucleotídeos. O termo *indel* é utilizado genericamente para designar inserção ou deleção quando, ao se comparar duas os mais sequências, não se sabe a origem do evento (Lemey, 2009).

Indels são raros em regiões codificantes e mais comuns em regiões não codificantes. Quando ocorrem em regiões codificantes, *indels* podem, ocasionalmente, modificar a matriz de leitura aberta do gene tornando outra matriz de leitura acessível, o que pode levar o gene a adquirir uma nova função. Por possuírem genomas pequenos e por não possuírem

introns (portanto, sendo incapazes de fazer splicing alternativo), vírus e bactérias lançam mão deste recurso amplamente para poder codificar diferentes proteínas à partir de um único gene. Este fenômeno é conhecido como Sobreposição de ORFs (*Overlapping ORFs*) (Lemey, 2009).

Mutações em um gene que são passadas adiante para as gerações descendentes e que coexistem entre si e com o estado ancestral do gene são denomidadas polimorfismos. Desta forma, para um determinado sítio polimórfico, variantes do mesmo gene circulam numa população simultaneamente (Lemey, 2009).

Optamos por sequenciar apenas a ORF do gene, ou seja, a sua porção codificadora. Sabe-se que as regiões reguladoras e codificadoras do gene estão submetidas a regimes distintos de pressões seletivas, estando fora do objetivo deste trabalho avaliar como a seleção natural está agindo a nível das regiões reguladoras do gene em questão. A ocorrência de mutações de ponto em regiões não codificantes está sujeita a outras imposições e restrições, tais como conservação dos sítios de ligação para fatores de transcrição, polimerases, RNA transportador, etc (Lemey, 2009).

A análise baseou-se tão somente na região codificante do gene, porque queremos verificar o efeito do fenótipo (proteína) por ela produzido e o impacto da seleção natural sobre este fenótipo, bem como suas repercussões sobre a adaptabilidade da bactéria ao seu meio. e como ela lida com a pressão seletiva exercida pelo sistema imune do hospedeiro.

Primeiramente foi gerado um alinhamento das sequencias de nucleotídeos correspondentes ao gene *sodC* de *N. meningitidis*. Existem muitas sequências deste gene para esta espéciedepositadas no Genebank, porém, na grande maioria dos casos, o gene não foi sequenciado por inteiro, de maneira que as sequências que constam neste banco de

dados exibem tamanhos diferentes, ou diferentes regiões do gene foram sequenciadas com propósitos distintos.

Numa primeira etapa, com a finalidade de não perder o poder do alinhamento, optamos por fazê-lo utilizando somente as 15 sequências obtidas à partir do prórpio laboratório, já que todas possuem exatamente 560pb, que corresponde ao tamanho exato da ORF do gene *sodC* em *N. meningitidis* (constatação feita à partir de sequências. provenientes de genomas completos depositados em bancos de dados públicos).À partir deste alinhamento uma árvore filogenética foi gerada (figura 19). Inicialmente optamos por não enraizar a árvore.

Nesta árvore, pode-se observar a presença de três ramos principais (ou maiores) distintos. O primeiro (destacado em vermelho), e também o maior deles, agrupa as sequências pertencentes às linhagens dos sorogrupos B e C, que são os sorogrupos com

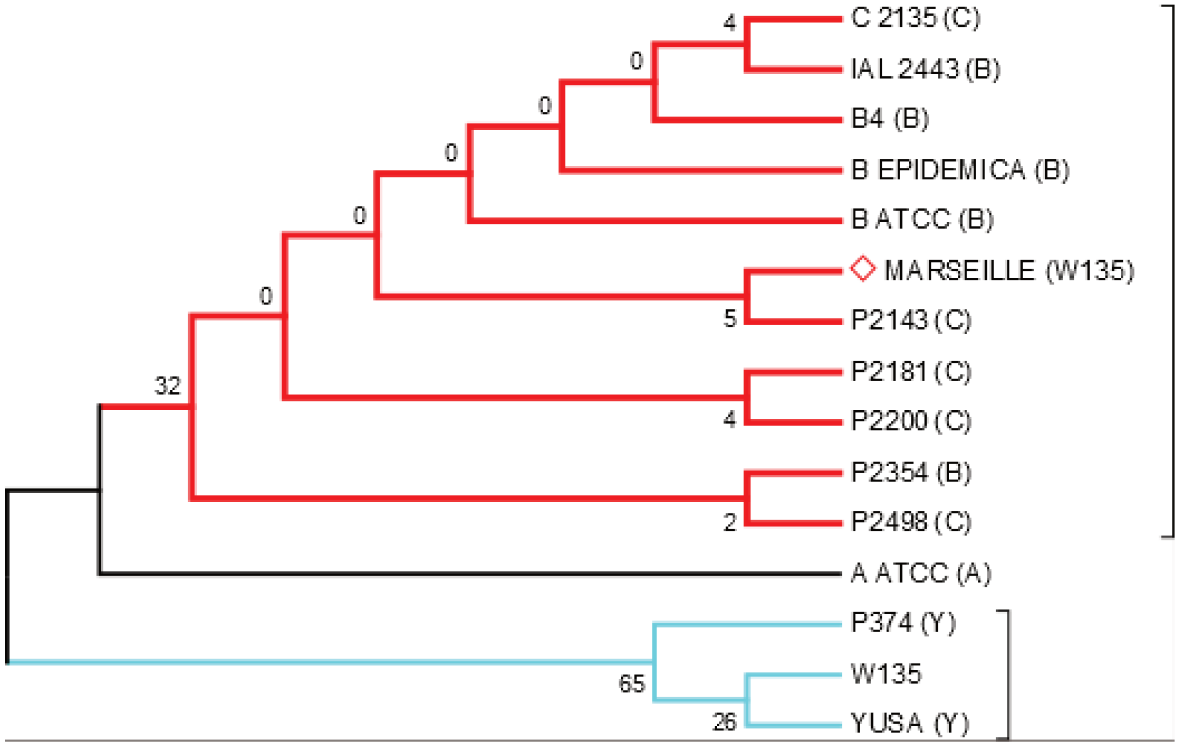


Figura 19. Árvore 1 para o gene *sodC* de *N. meningitidis* : representação das 15 amostras pertencentes ao LABIOTEC.

maior relevância epidemiológica, por possuirem distribuição mundial e também por serem os principais cauradores (90% dos casos) de surtos epidêmicos de doença muito grave, com choque séptico e morte. Para o sorogrupo C já existe vacina comercial disponível, porém são muitos os relatos de escape vacinal. Para o sorogrupo B não existe vacina disponível,

dado que este sorogrupo possui na sua superfície moléculas que mimetizam as moléculas do hospedeiro, portanto a resposta imune a este sorogrupo é muito fraca (Stephens, 2007).

O terceiro ramo, destacado em azul, agrupa as linhagens pertencentes aos sorogrupos Y e W135, com exceção da linhagem denominada MARSEILLE , destacada por um losango vermelho, que pertence ao sorogrupo W135, mas que agrupa junto com as linhagens pertencentes aos sorogrupos B e C. Foi relatado que esta linhagem foi responsável por um surto de doença meningocócica devastadora na cidade de Marseille, interior da França, o que é atípico para as linhagens do sorogrupo W135 (Taha, 2000). Portanto, esta linhagem pode ter sido tipada de forma errônea, ou seu gene *sodC*pode se comportar de forma semelhante aos genes dos sorogrupos B e C, o que nos permite questionar se o gene *sodC*, especificamente desta linhagem, foi adquirido por transferência lateral e recombinação homóloga à partir de uma outra linhagem do sorogrupo B ou C, tornando-a mais virulenta.

Para uma estimativa realística das relações filogenéticas entre genes, toda a sequência sob investigação deve exibir a mesma história evolutiva. Eventos de recombinação homóloga dentro do gene afetarão a inferência filogenética, pois porções de sequências distintas serão combinadas dentro do mesmo gene. No entanto, a recombinação fora do fragmento de interesse não altera a topologia da árvore. (Lemey, 2009).

Em relação às demais linhagens deste ramo, sabe-se que os sorogrupos Y e W135 são menos virulentos, porém têm importância epidemiológica por apresentarem distribuição mundial e por causarem doença em crianças, idosos e agrupamentos humanos (recrutas militares, alojamentos estudantis, etc) (Yasdankhah, 2004).

O segundo ramo, destacado em preto, agrupa a única linhagem pertencente ao sorogrupo A, grupo virulento e causador de doença grave, porém restrito à África Subsaariana, não apresentando,portanto, distribuição mundial (Stephens, 2007).

Ao analisar o alinhamento que gerou esta árvore (Figura 20), constatamos que as sequências se mostraram muito conservadas, ou seja, quase nenhuma variabilidade (na forma de mutação) foi detectada, com exceção de duas substituições não sinônimas e nenhuma substituição sinônima, o que é bastante incomum para sequências que, assume-se, já estão divergindo há algum tempo ao longo da sua história evolutiva.

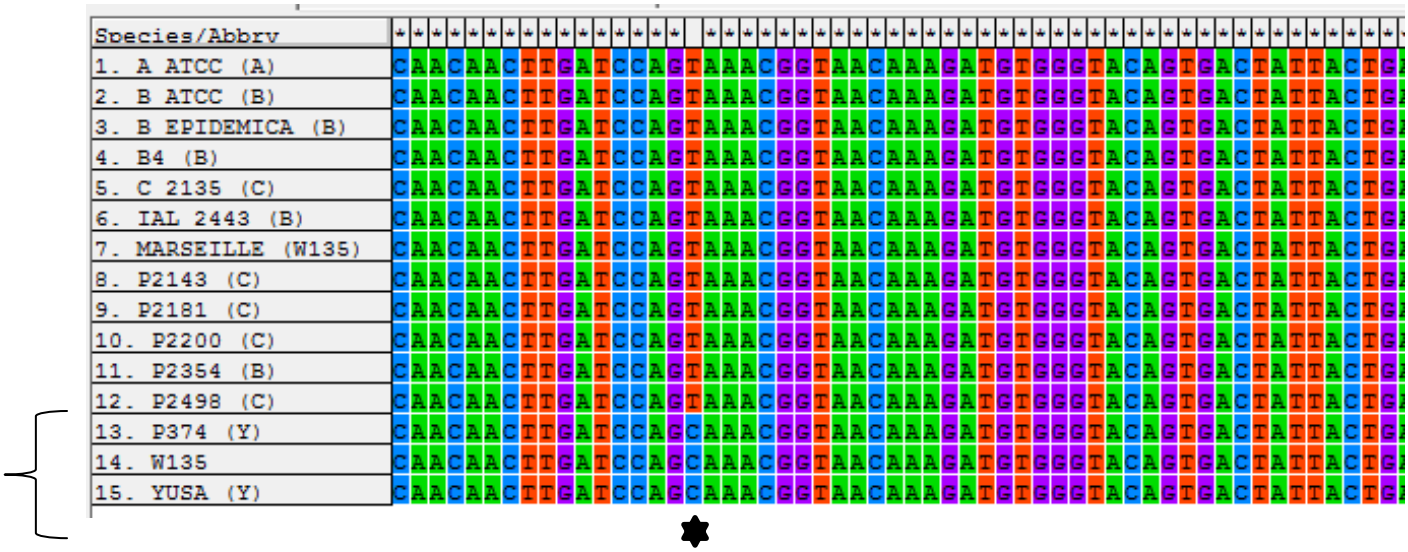


Figura 20: Alinhamento parcial de 15 sequências de nucleotídeos correspondentes ao gene *sodC*de *N. meningitidis*. Na figura, encontra-se destacada por uma estrela preta a substituição sinônima que definiu o ramo que agrupa as sequências pertencentes aos sorogrupos Y e W135 (entre parênteses). Para uma melhor apreciação de todas as substituições de nucleotídeos deste alinhamento, ver Tabela 9.

Substituições não sinônimas correspondem à modificação de um nucleotídeo no codon que acarreta na modificação do aminoácido correspondente na proteína. Ao passo que, substituições sinônimas correspondem à modificação de nucleotídeo sem mudança do aminoácido correspondente, porque o código genético é degenerado (Yang, 1998).

Portanto, substituições sinônimas são ditas silenciosas, de forma que sequências que estão divergindo há algum tempo tendem a acumular mais rapidamente substituições sinônimas (justamente por serem silenciosas, não levama modificações na sequência e estrutura terciária da proteína), enquanto que o acúmulo de substituições não sinônimas está na dependência do tipo de força seletiva que está atuando sobre as sequências. Se uma restrição funcional muito forte é imposta a uma proteína, isto significa que ela apresenta um papel primordial (quer seja numa rota metabólica ou como receptor de membrana, por exemplo), de forma que a sua sequência e, principalmente, a sua estrutura terciária não podem variar (o que poderia acarretar em inativação do sítio ativo, por exemplo). Neste caso, o gene que codifica esta proteína estará sob ação da seleção negativa ou purificadora, que é aquela que tende a manter a sequência inaltarada ou com poucas alterações, para não comprometer o *folding* e o funcionamento da proteína. Consequentemente, as novidades adaptativas (ou seja, mutações) tendem a ser eliminadas da sequência (Balding, 2007).

No outro caso de força determinística, que é a seleção positiva ou adaptativa, novidades evolutivas ou adaptativas (mutações) são aceitas ao longo da sequência por conferir à proteína por ela codificada algum tipo de vantagem em relação ao meio. Por exemplo, se considerarmos o caso de proteínas de interface com o sitema imune, é vantajoso o acúmulo de modificações que permitam o não reconhecimento do patógeno e o consequente escape do sistema imune do hospedeiro (Balding, 2007).

Vale ressaltar que as notas de *bootstrap* (reamostragem com reposição) atribuídas a esta árvore são muito baixas, não oferecendo, portanto, suporte para asua topologia.

Portanto, devido a ausência de mutações entre as sequências e a baixa significância estatística da árvore, optamos por adicionar mais 15 sequências do Genebank (Figura 22) com o mesmo comprimento de 560 pb no intuito de acrescentar maior variabilidade ao

alinhamento. Uma nova árvore filogenética (Figura 21) foi reconstruída com o objetivo de observar se o padrão dos ramos se manteriao mesmo ou se sofreria alterações.

Foi constatado que este acréscimo, todavia, não acarretou em alterações no padrão da topologia da árvore filogenética do gene *sodC* em *N. meningitidis,* nem introduziu variabilidade ao alinhamento, e o mais importante: não apareceram substituições sinônimas. Foi encontrada apenas uma terceira substituição não sinônima, totalizando 3 substituições não sinônimas e nenhuma substituição sinônima.

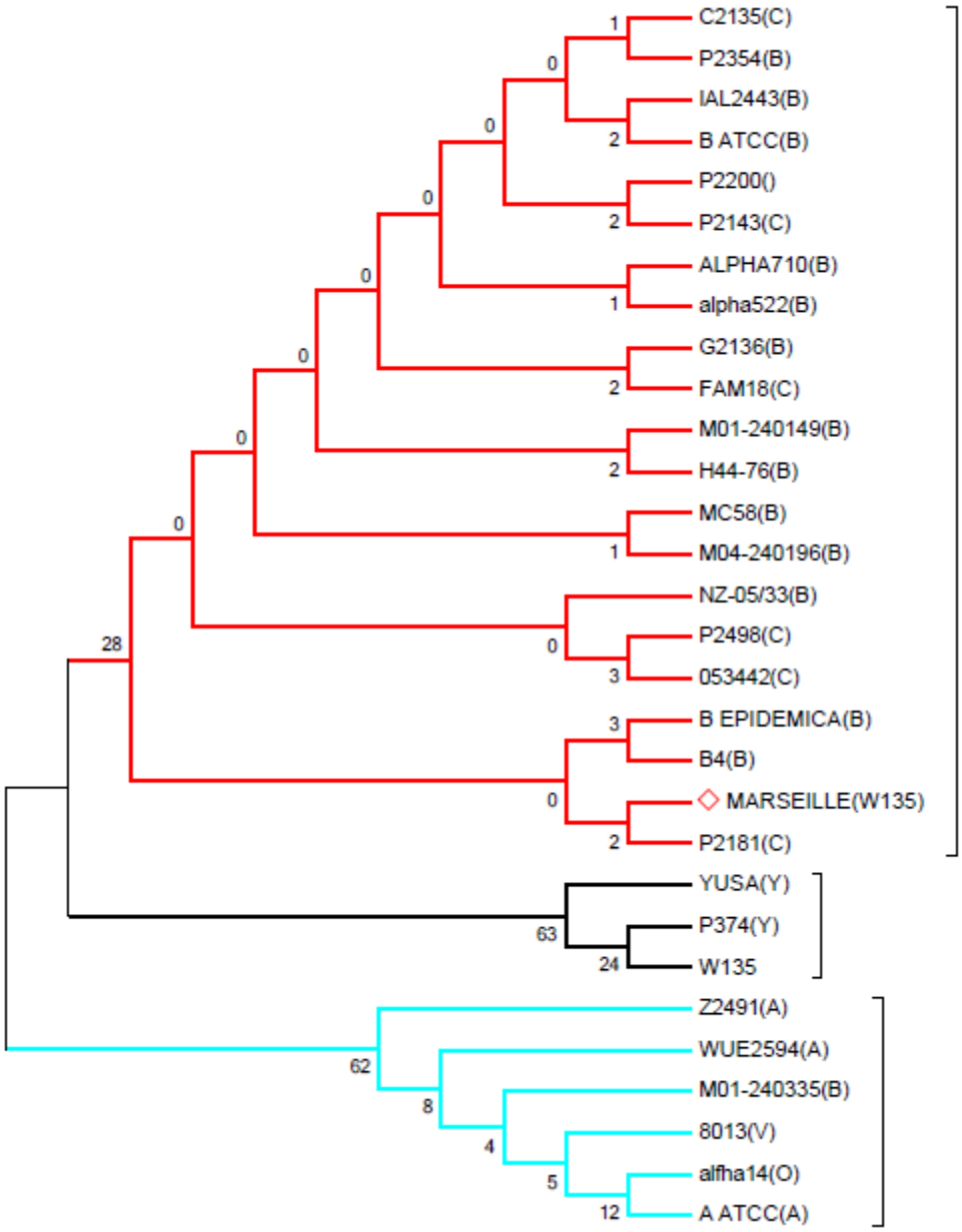


Figura 21. Árvore 2 para o gene *sodC* de *N. meningitidis* :o acréscimode 15 sequências do Genebank não acarretou na modificação da topologia da árvore.

A segunda árvore reconstruída, portanto, continuou exibindo três ramos, sendo que o primeiro, destacado em vermelho, agrupa as linhagens pertencentes aos sorogrupos B e C, juntamente com a linhagem MARSEILLE (sorogrupo W135 – destacada por um losango vermelho). Neste ramo especificamente, não foram observadas alterações de nucleotídeos

(as sequências são idênticas). Uma primeira substituição não sinônima deu origem ao segundo ramo que agrupa as linhagens pertencentes aos sorogrupos Y e W135, destacado em preto. A segunda substituição não sinônima deu origem ao terceiro ramo, que agrupa principalmente linhagens pertencentes ao sorogrupo A, destacado em azul, e a terceira mutação constitui um *singleton,*ou seja, ocorreu isoladamente na linhagem alpha14_O, dentro do terceiro ramo. Esta linhagem pertence ao sorogrupo O, o qual não tem relevância epidemiológica (Stephens, 2007).

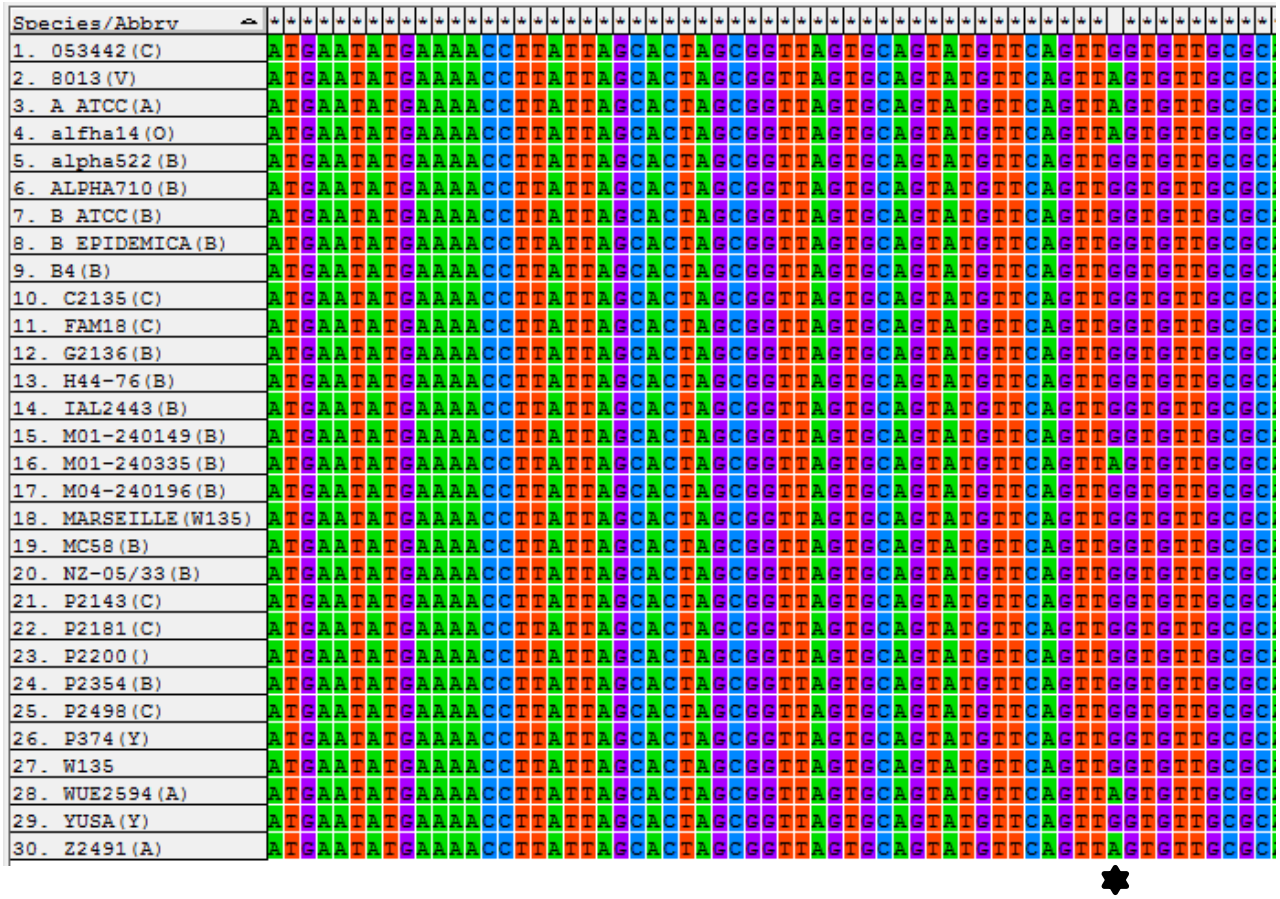


Figura 22: Alinhamento 2, com 30 sequências do gene *sodC*de *N. meningitidis.* Destacado por uma estrela em preto,observa-se uma mutação encontrada preferencialmente em linhagens pertencentes ao sorogrupo A. Para uma melhor apreciação das sibstituições encontradas neste alinhamento, ver Tabela 9.

Em relação às substituições de aminoácidos encontradas, constatou-se o seguinte (Figura 16) : em uma sequência de nucleotídeos de 560pb, que codifica uma enzima com 186 resíduos de aminoácidos, sendo que os 22 primeiros aminoácios correspondem ao peptídeo sinal, observou-se (tabela 9):

1) No sítio 52 do nucleotídeo, que corresponde ao sítio 18 da enzima (portanto, a nível do peptídeo-sinal), um **G** foi substituído por um **A,** fazendo com que o codon **GGT** fosse trocado por um **AGT.** Neste caso, uma **GLICINA** foi substituída por uma **SERINA**, que são aminoácidos de classes diferentes. Este evento, cujo padrão se repetiu em algumas sequências, deu origem ao terceiro ramo da árvore (agrupando as linhagens do sorogrupo A principalmente, dentre outras).
2) No sítio 131 do nucleotídeo, que corresponde ao sítio 44 da proteína, um **T** foi substituído por um **C,** fazendo com que o codon **GTA** fosse substituído por um **GCA.** Neste caso, uma **VALINA** foi substituída por uma **ALANINA,** que são aminoácidos pertencentes à mesma classe funcional. Esta alteração deu origem ao ramo que aqrupa as linhagens dos sorogrupos Y e W135.
3) No sítio 304 do nucleotídeo, que corresponde ao sítio 102 da proteína, um **G** foi substituído por um **A,** fazendo com que um codon **GGC** fosse substituído por um **AGC.** Neste caso, uma **GLICINA** foi novamentesubstituída por uma **SERINA**, que são aminoácidos de classes

funcionais diferentes. Este evento ocorreu em uma linhagem isolada dentro do terceiro ramo, a alpha14_O.

Tabela 13: Substituições não sinônimas entre as sequências do gene *sodC* de *N. meningitidis.*

POSIÇÃO NUCLEOTÍDEO/PROTEÍNA		TROCA DO NUCLEOTÍDEO	TROCA DO CODON	TROCA DO AMINOÁCIDO	SOROGRUPO
52 (PEPTÍDEO SINAL)	18	G → A	GTT → AGT	GLICINA→SERINA	A
131	44	T → C	GTA → GCA	VALINA→ALANINA	Y, W135
304	102	G → A	GGC → AGC	GLICINA→SERINA	O

É digno de nota que serina, tirosina e treonina são aminoácidos com grande propensão a serem fosforilados e desfosforilados, tamando parte, desta forma, das rotas de sinalização intracelular. Observamos que, por duas vezes, glicina foi substituída por serina. Glicina é um aminoácido apolar, enquanto que serina é um aminoácido polar neutro, pertencendo, portanto, a classes diferentes. Dado que serina aceita facilmente um grupo fosfato e este é fortemente carregado negativamente, isto poderia alterar drasticamente os sítios onde ocorreram tais mudanças, visto que estes sítios eram hidrofóbicos anteriormente. Estas modificações podem ter sérias repercussões no *folding* da proteína.

Pelo fato de não havermos constatado que variabilidade foi introduzida ao segundo alinhamento e que isto não é compatível com sequências que estão divergindo há algum tempo, as seguintes hipóteses foram formuladas: 1) estas sequências estão divergindo há muito pouco tempo e variabilidade ainda não pode ser acumulada, portanto o evento de transferência lateral do gene *sodC*do gênero *Haemophilus* para a espécie *N. meningitidis* é muito recente na escala do tempo; ou 2) este gene, em *N. meningitidis* é muito conservado

devido a uma imposição ou restrição funcional muito forte. Esta última hipótese não é muito compatível com o comportamento de genes que codificam proteínas que fazem parte da interface entre o sistema imune do hospedeiro e o parasito, visto que, para escapar do reconhecimento do sistema imune, estes genes acumulam variabilidade (ou seja, mutações) muito rapidamente. Esta hipótese somente seria rejeitada se o gene da SodC efetivamente não fizesse parte desta interface com o sistema imune, ou seja, se efetivamente a proteína exercesse outra função (exibisse um outro fenótipo) que não a que foi previamente descrita e postulada pela literatura (Kroll, 1998; Dunn, 2003).

Vale ressaltar que mesmo com 15 sequências a mais, as notas do teste de confiabilidade amostral ainda são muio baixas, o que não sustenta a topologia desta nova árvore. Este viés pode estar relacionado a um número insuficiente de amostras. Desta forma, como a amostragem do segundo alinhamento ainda estava pequena e a pouquíssima variabilidade observada entre as sequências ainda não corroborava a hipótese mais plausível (de que as sequências já estavam divergindo há algum tempo, e,portanto, deveria ter acontecido o acúmulo de um número maior de mutações, principlamente na forma de substituições sinônimas), todas as sequências disponíveis no Genebank para a SodC de *N. meningitidis* foram utilizadas para fazer um novo alinhamento, totalizando 50 sequências (Figura 23). Como as sequências possuiam tamanhos diferentes, porque alguns genes não foram sequenciados por inteiro, um alinhamento foi feito e, posteriormente, as porções com tamanhos diferentes nas extremidades de cada sequência foram removidas manualmente, para que todas as sequências tivessem o mesmo tamnho final e penalizações geradas pelo programa computacional não fossem introduzidas erroneamente. Isto resultou na diminuição do tamanho final de cada sequência de 560pb para 429pb. Como alguns genes foram sequenciados sem a sua porção inicial, um dos parâmetros para a reconstrução das

filogenias anteriores foi perdido : a transição de uma guanina para uma adenina, gerando a substituição de uma glicina por uma serina. Este parâmetro definia o ramo que agrupava as linhagens pertencentes ao sorogrupo A, e foi perdido neste novo alinhamento. Em contrapartida, novas mutações foram introduzidas: mais uma substituição não sinônima e duas substituições sinônimas. Uma nova filogenia foi reconstruída e a sua topologia foi analisada (Figura 24).

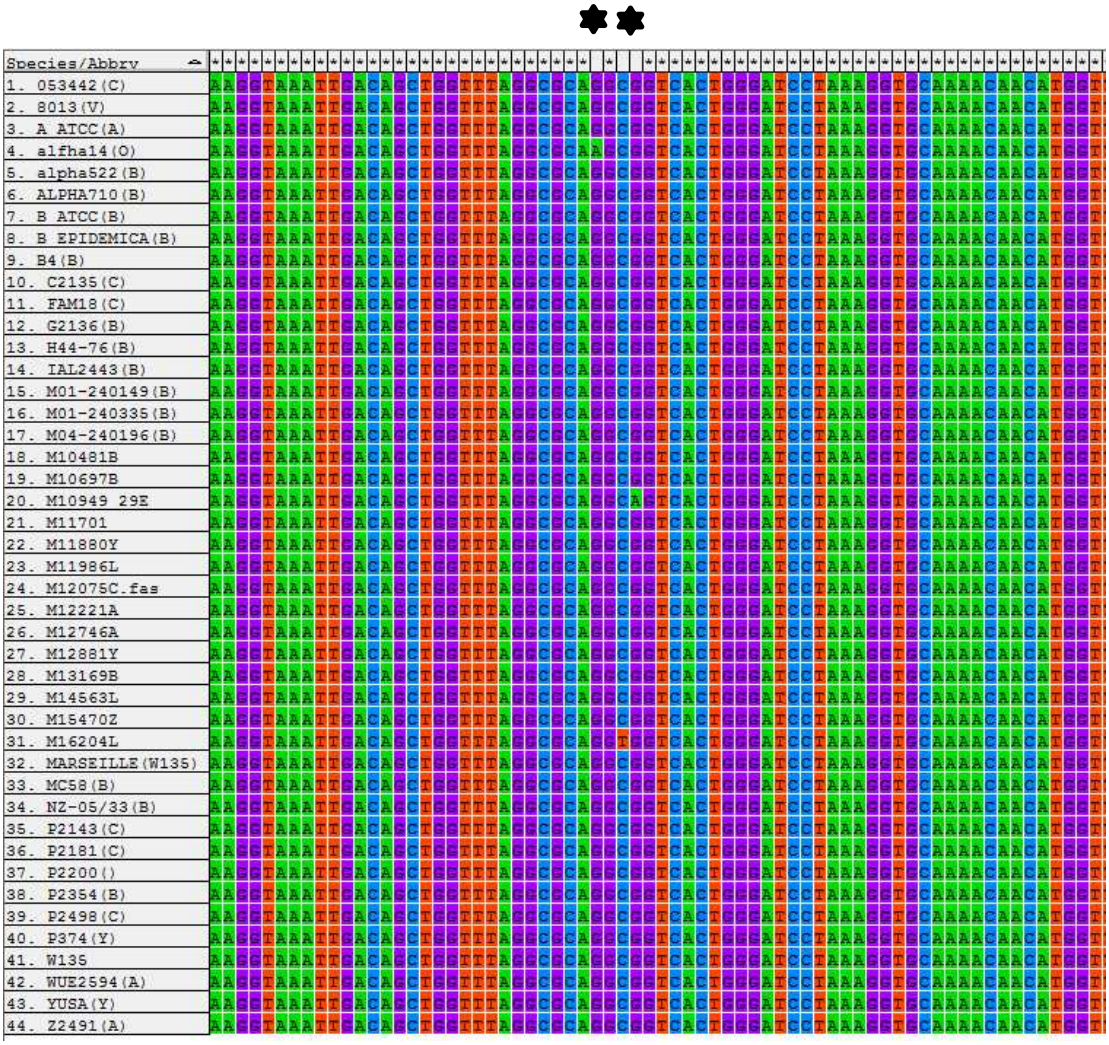


Figura 23: Alinhamento 3, contendo todas as sequências disponíveis para o gene *sodC* de *N. meningitidis.* As estrelas destacam os *singletons* encontrados neste alinhamento.

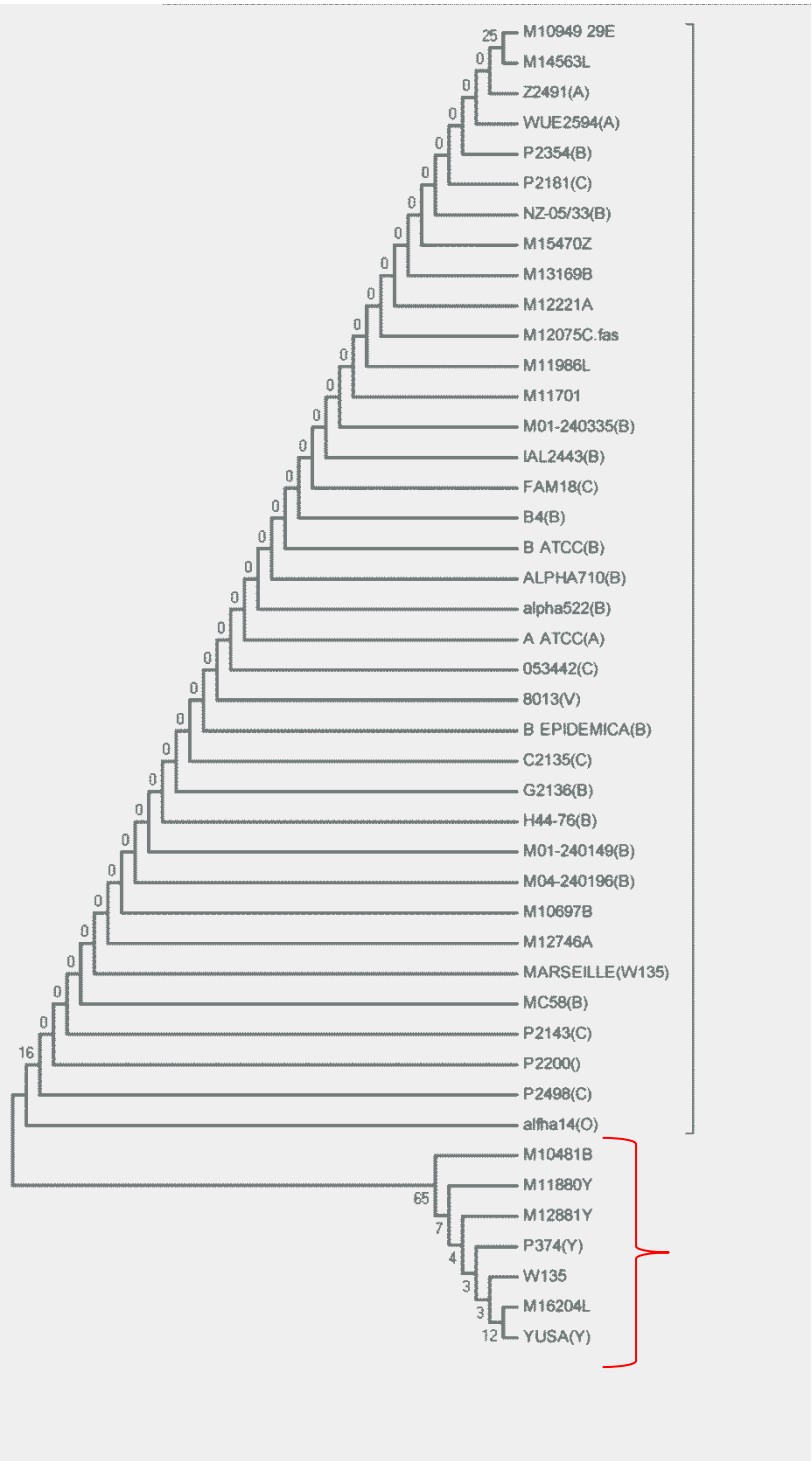


Figura 24: Árvore 3 para o gene *sodC* de *N. meningitidis* com todas as sequências disponíveis no Genebank mais as sequências do LABIOTEC. Notar perda do ramo que agrupava sequências pertencentes ao Sorogrupo A.

Nesta nova árvore, constatamos a presença de dois ramos principais. O ramo que agrupa predominantemente as linhagens dos sosrogrupos Y e W135 está destacado por um parêntese vermelho na figura.Observa-se, ainda, que novas mutações, duas sinônimas e uma não sinônima, ocorreram, porém, em sequências isoladas, constituindo *singletons*(Figura 17). Estas mutações isoladas, na maioria das vezes, constituem formas deletérias e, com o tempo, são eliminadas por seleção negativa (aquele que previne a ocorrência de má adaptações a nível da proteína). Caso sejam vantajosas, ainda não houve tempo o suficiente para que elas fossem mantidas por seleção positiva (a que permite a inclusão de novidades adaptativas à sequência, quando essas novidades são benéficas ou conferem uma aptidão maior ao organismo em questão em relação ao seu meio), e passadas adiante para constituir um padrão de distribuição ao longo do alinhamento.Esses padrões de distribuição de mutações ao longo alinhamento têm suma importância, pois são eles que constituem os parâmetros através dos quais topologias de ramos são geradas em árvores filogenéticas, bem como suposições de homologia são formuladas.(Lemey, 2009)

Tendo em vista o que já foi exposto anteriormente, parece-nos que o evento de transferência lateral de genes, através do qual a espécie *N. meningitidis* adquiriu e fixou o gene *sodC* aconteceu num passado muito recente, não tendo havido tempo habil o suficiente para que variabilidade genética fosse adquirida entre as sequências dentro da mesma espécie, principalmente através do acúmulo de substituições sinônimas, as quais tendem a ser silenciosas e, portanto, não são varridas da sequência. Além disto, nos testes de confiabilidade amostral (*bootstrap* – reamostragem com reposição), as notas obtidas pelos ramos foram muito baixas (inferior a 85%), o que quer dizer que esta topologia de árvore não é confiável.Isto quer dizer que é mais provável que as sequências se agrupem num ramo só do que em três ramos distintos. Isto não invalida o que foi observado no

alinhamento, apenas reforça a hipótese de que o evento de transferância lateral foi muito recente e mutações ainda não foram acumuladas ao longo das sequências, pelo menos não o suficiente para gerar um padrão de distribuição que permita reconstruir uma árvore com uma topologia confiável.

Portanto, observa-se um evento muito recente de aquisição de um novo fator de virulência. Por conferir um fenótipo vantajoso ao patógeno, permitindo o escape da fagocitose e pelo fato do gene estar amplamente presente em todas as linhagens analisadas, supõem-se que este gene esteja se fixando nesta espécie pela ação de algum de força determinística (seleção natural – positiva ou negativa) e não por ação de algum processo estocástico (deriva genética aleatória).

Para averiguar o fenótipo conferido por cada mutação à sequência da proteína, se vantajoso ou deletério, bem como suas implicações no folding e na estrutura terciária, será necessária a realização de modelagem proteica de suas sequências, cada uma contendo uma das duas substituições não sinônimas que apresentaram um padrão de distribuição ao longo do alinhamento, e que foram, portanto, determinantes para o surgimento de dois dos três ramos da árvore filogenética para o gene *sodC* de *N. meningitidis.*

5.2 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC* do gênero *Haemophilus.*

Das 45 linhagens de *Haemophilus influenzae*pertencentes ao LABIOTEC, o gene *sodC* foi amplificado por PCR nas condições-padrão previamente descritas em apenas 3 linhagens: Hi 12 e Hi 19, que são linhagens não tipáveis – NTHi; e Hi E, uma linhagem de *H. influenzae* capsulado) (Moxon, 1985). Este achado corrobora com o que já foi descrito

na literatura: o gene *sodC* está presente apenas em 17% das linhagens de *H. influenzae.*, parecendo ser um gene escasso entre as linhagens desta espécie e que, por algum motivo, não atingiu a fixação como observado em outras espécies bacterianas como, por exemplo, a espécie *N. meningitidis* (Kroll, 1991; McCrea, 2010).

Desta forma, para a reconstrução da filogenia do gênero *Haemophilus*foram utilizadas predominantemente sequências disponíveisno Genebank para a espécie*H. influenzae.* Também foram acrescidas sequências pertencentes às espécies *H. parainfluenzae, H. haemolyticus e H. ducrey.* Para o enraizamento da árvore foi utilizada uma sequência de *Yersinia pestis*.Assim, um alinhamento foi gerado e uma filogenia foi estimada para este gênero bacteriano (Figura 25).

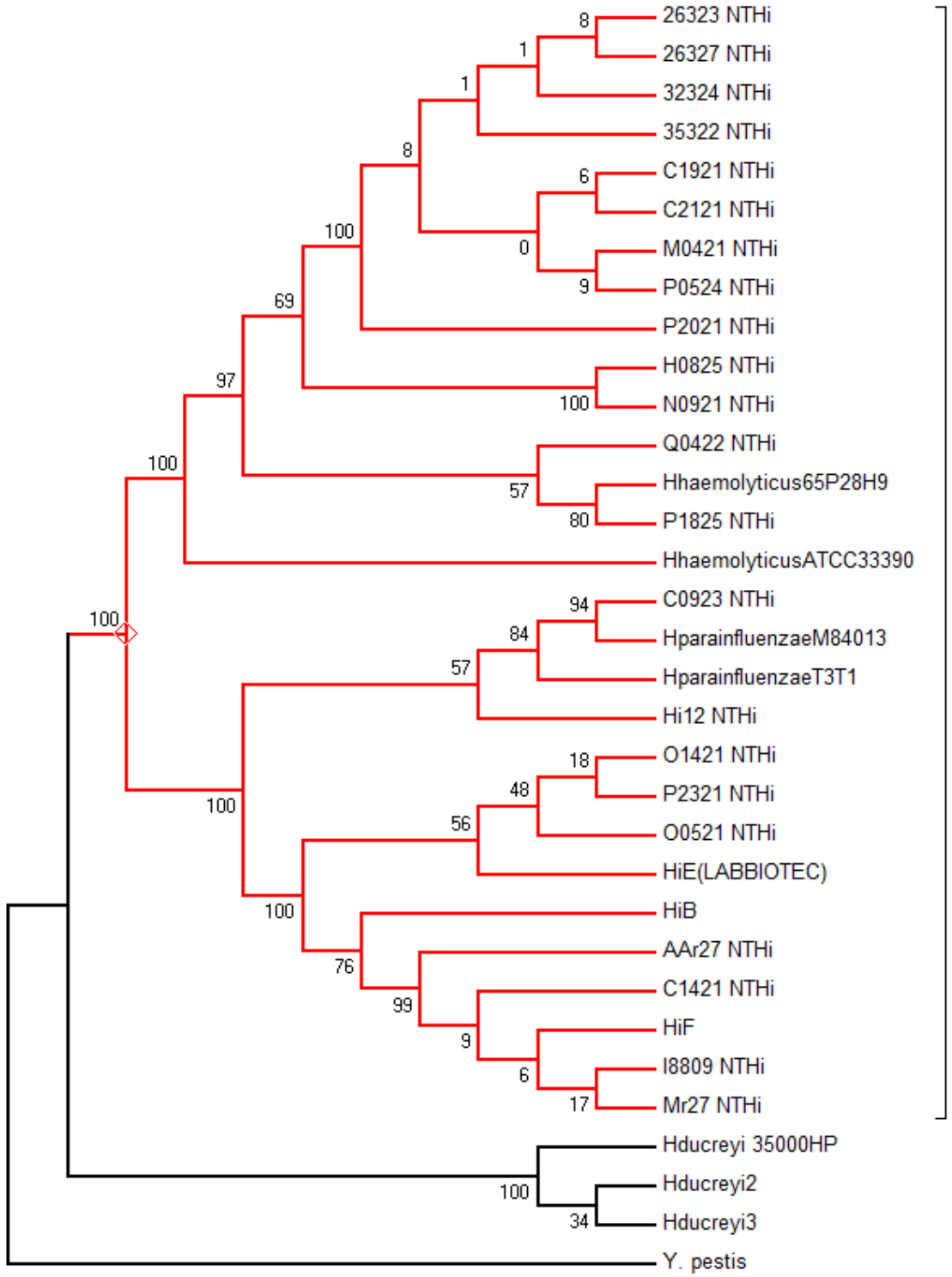


Figura 25: Árvore filogenética para o gene *sodC* de espécies do gênero *Haemophilus*.

Ao se observar esta árvore, podemos constatar a presença de dois ramos externos distintos. Nota-se que as notas do teste de confiabilidade amostral são altas o bastante para sustentar esta topologia de árvore. Um dos ramos, destacado na figura em preto,agrupa as sequências pertencentes à espécie *H. ducrey*. A proteína SodC desta espécie é a única descrita até o momento que exibe a peculiaridade de ligar o grupo prostético Heme (Chilemi, 2012). Portanto, a sequência do gene correspondente deve ter acumulado mutações ou novidades adaptativas para permitir esta singularidade, de forma que, todas as sequêncas de *H. ducrey*, por compartilharem a mesma novidade evolutiva que os distingue das demais, agrupam-se num ramo específico.

O outro ramo, destacado em vermelho, agrupa as demais espécies do gênero em questão. Este ramo se bifurca em dois outros ramos internos, que também exibem notas de bootstrap bastante altas, o que torna esta divisão (topologia) bastante confiável. Num deles, destacado em rosa (Figura 26), algumas sequências de *H. influenzae* não tipáveis (NTHi) compartilham um ancestral comum com a espécie *H. haemolyticus.* No outro ramo, destacado em azul, as demais sequências de *H. influenzae* não tipáveis bem como todas as linhagens capsuladas de *H. influenzae* (marcadoas com um losango azul) compartilham um ancestral comum com a espécie *H. parainfluenzae*. Ao se realizar o teste de confiabilidade amostral (*Bootstrap* – reamostragem com reposição), observa-se que a nota destes ramos são bastantes altas (todas superiores a 85%), de forma que a topologia dos mesmos é bastante confiável.

Nota-se que uma parte das sequências das linhagens não tipáveis de *H. influenzae* exibe alta similaridade com o gene *sodC* da espécie *H. haemolyticus.* Como eventos de transferência lateral de genes são bastante frequentes em toda a família *Pasteurellaceae* (Findlay, 2009), nós hipotetizamos que uma parte das linhagens de *H. influenzae* recebeu o

gene *sodC* por transferência lateral da espécie *H. haemolyticus*, visto que o score de similaridade par a par destas sequências com a sequência de *H. haemolyticus* é muito alto, e muito menor em relação à espécie *H. parainfluenzae*, a qual agrupa no ramo parálogo a este.

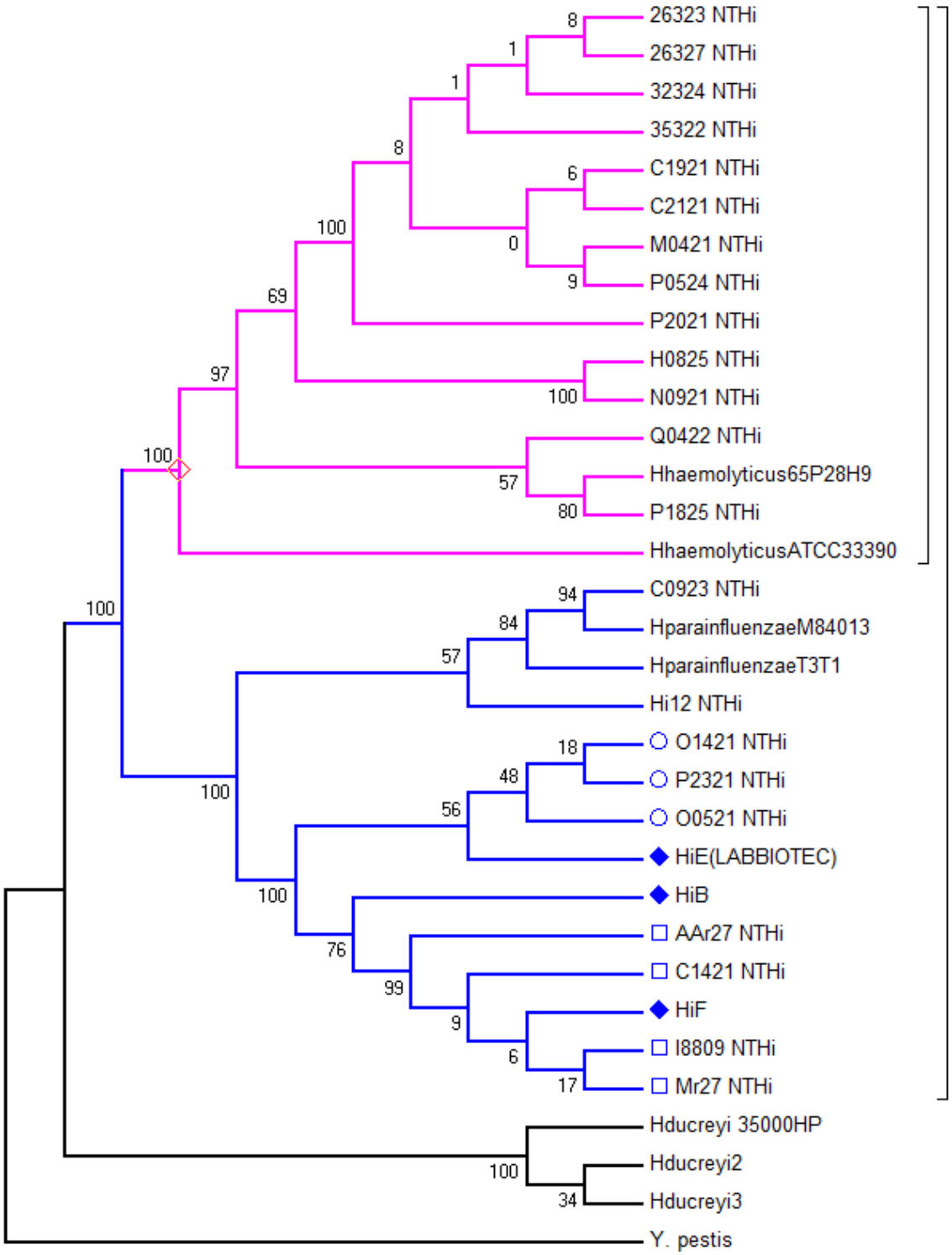


Figura 26: mesma árvore anterior, destancando os ramos que compartilham um ancestral comum com *H. haemolyticus* e com *H. parainfluenzae*.

A outra parte das sequências de *H. influenzae* não tipáveis, juntamente com todas as sequências de *H. influenzae* capsuladassão bastante semelhantes à sequência de *H. parainfluenzae*, visto que o score de similaridade par a par entre estas e a sequência de *H. parainflueanzae* é muito alto, o que não ocorre em relação à sequência de *H. haemolyticus*. Além disso, à partir da observação da topologia da árvore, nota-se que *H. parainfluenzae* compartilha um ancestral comum com estas sequências, que é distinto do ancestral comum compartilhado entre *H. haemolyticus* e as demais sequências de *H. influenzae.*

À partir desta observação, nós concluimos que estas linhagens de *H.influenzae* receberam o gene da SodC por transferência lateral da espécie *H. parainfluenzae.* Visto que a nota destes ramos da árvore é alta o suficiente para apoiar esta hipótese, e que eventos de transferência lateral de genes são muito frequentes na família *Pasteurellaceae*, foi constatado o seguinte: uma única espécie, *H. influenzae*, recebeu o mesmo gene, *sodC*, de duas fontes distintas: uma, o *H. haemolyticus*, o que já havia sido apontado em estudos anteriores (McCrea, 2010), e a outra o *H. parainfluenzae.*

Ao se analisar mais apuradamente as sequências de nucleotídeos deste alinhamento (Figuras 27 e 28), observamos que as sequências mais aparentadas com *H. haemolyticus*constituem um gene verdadeiro, com matriz de leitura aberta íntegra. Em contrapartida, todas as sequências mais aparentadas com *H. parainfluenzae* (marcadas com um círculo ee com um quadrado azul), com exceção das sequências de *H. influenzae* não tipáveis Hi 12 e C0923, e das sequências de *H. influenzae* capsuladas HiF e HiB, sofreram um ou mais eventos de pseudogeneização visto que todas elas exibem mais de um stop códon prematuro ao longo de sua matriz de leitura.

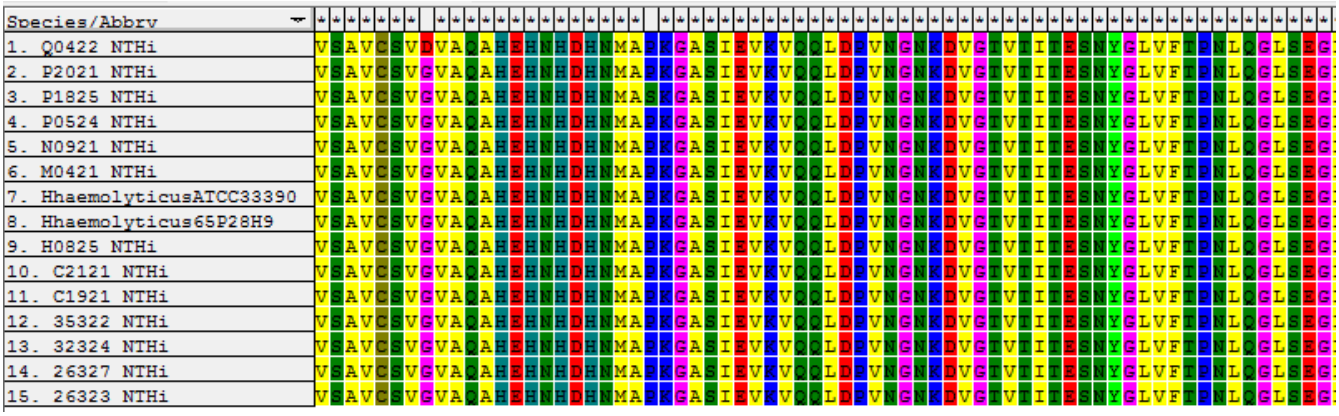


Figura 27: Alinhamento parcial das sequências que compartilham um ancestral comum com *H. haemolyticus*. As sequências de nucleotídeos foram traduzidas computacionalmente, não se observando *stop códons* prematuros ao longo da sequência.

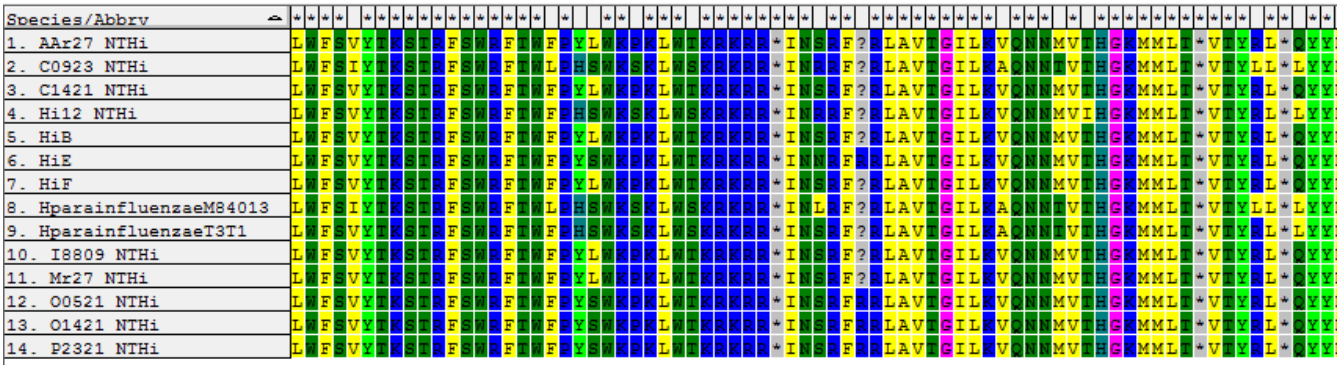


Figura 28: Alinhamento parcial das sequências que compartilham um ancestral comum com H. parainfluenzae. As sequências de nucleotídeos fora, traduzidas computacionalmente. As colunas em cinza representam stop códons prematuros ao longo da sequência.

Se as mesmas sequências compartilham os mesmos stop códons nas mesmas posições, é provável que um ou mais eventos de transferência lateral, que sucederam o evento inicial, tenham ocorrido dentro do gene de uma linhagem ancestral, tirando a matriz de leitura do frame e gerando um pseudogene. Esta característica foi, então, passada adiante para as linhagens descendentes. Verificamos que as sequências Mr27, AAr27, I8809 e C1421 (que compõem o mesmo ramo, juntamente com as sequências capsuladas HiB e

HiF, e que estão destacadas por um quadrado azul), possuem 5 stop códons prematuros idênticos ao longo da sua matriz, nas seguintes posições do alinhamento: 256, 298, 391, 415 e 469. Como estas sequências compartilham os mesmos stop códons nas mesmas posições, então nós concluimos que um único evento intragênico de transferência lateral ocorreuna posição 256, desalinhando o restante da matriz de leitura do gene e originando os demais stop códons.

Para uma análise mais minunciosa desta situação, estas sequências foram comparadas com as duas sequências de *H. parainfluenzae*e com uma sequência de *H. influenzae* que possui um gene íntegro (26323_NTHi) (Figura 29): em relação à sequência 26323_NTHi, tanto as sequências de *H. parainfluenzae* quanto as sequências AAr27, Mr27, I8809 e C1421 possuem uma deleção de 7pb na posição 47 do alinhamento de nucleotídeos. Por não ser múltipla de 3, esta deleção tiraria a matriz do frame, o que tornaria as sequências de *H. parainfluenzae* pseudogenes também. No entanto, observa-se, logo em seguida, na posição 62 do alinhamento, outra deleção, desta vez de 5pb, que também não é múltipla de 3, porém se somada a anterior totaliza 12pb que é múltiplo de 3.Portanto esta última deleção colocou a matriz de leitura de volta ao frame, o que corrobora o fato de que o gene *sodC*de *H. parainfluenzae* não é um pseudogene. Isto não impede, todavia, que o gene *sodC*de *H. parainfluenzae* tenha sido transmitido parcialmente a estas sequências (por transferência lateral intragênica de um pedaço do gene), gerando uma matriz truncada e não codificante. Ou, ainda, que outros *indels* tenham tirado a matriz destas sequências do frame.

Comparando-se as sequências AAr27, Mr27, I8809 e C1421 com a sequência de *H. parainfluenzae* T3T1, observou-se na posição 149 do alinhamento uma inserção de 7pb não compartilhada com a sequência de *H. parainfluenzae*. Portanto, este foi o evento que tirou a matriz de leitura aberta destas sequências do frame, resultando no aparecimento de 5 stop

códons prematuros. Esta inserção foi removida manualmente das sequências que, ao serem traduzidas computacionalmente, geraramuma proteína sem stop códons prematuros.

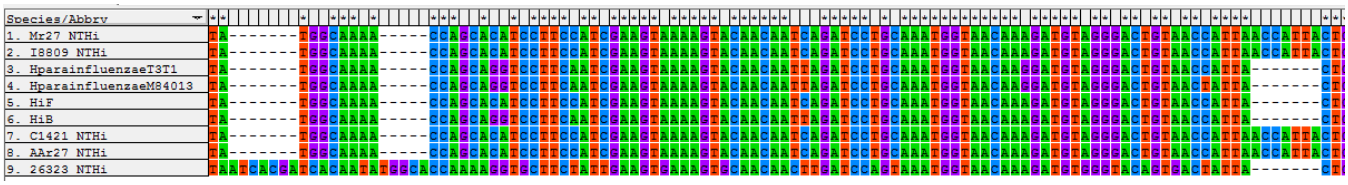


Figura 29: Alinhamento parcial das sequências AAr27, Mr27, I8809, C1421 (pseudogenes), HiB e HiF com as sequências de *H. parainfluenzae* e 2632_NTHi (genes verdadeiros). Notar a presença e a dinâmica de *indels* tirando e recolocando a matriz de leitura no frame.

Como foi dito anteriormente, as sequências de *H. influenzae* capsuladas presentes neste ramo, HiB e HiF não possuem esta última inserção e, ao serem traduzidas computacionalmente, parecem codificar uma proteína íntegra.Vale ressaltar que a presença de *indels* na ORF de um gene é um evento puramente aleatório e amplamente utilizado por procariotos para gerar variabilidade alélica. Como todo evento estocástico, mutações deletérias ocorrem com a mesma chance ou probabilidade que mutações silenciosas e mutações vantajosas (Lemey, 2009).

Ao analisarmos as sequências O0521, O1421, P2321 e HiE, que constituem o outro ramo do clado mais parecido com *H. parainfluenzae,* notamos o seguinte: estas sequências possuem 4 stop códons prematuros em sua matriz de leitura, fazendo com que pseudogenes fossem gerados. Estes stop códons prematuros ocorrem na mesma posição dentro destas quatro sequências (posições 292, 385, 409 e 463), porém estão em posições distintas em relação às sequências anteriores. Desta forma, eventos distintos geraram este segundo grupo de pseudogenes.

No intuito de tentar traçar a origem ou as origens deste evento (Figura 30), estas sequências foram comparadas com uma sequência de *H. parainfluenzae*(*H. parainfluenzae* T3T1), com um pseudogene distinto do grupo anterior (AAr27) e com uma sequência de *H. influenzae* íntegra (26323_NTHi). Foi constatado o seguinte: estas sequências são bastante parecidas com a sequência de *H. parainfluenzae,* pois possuem as duas primeiras deleções iniciais, que somadas não tiram a matriz do frame, além de não possuírem a deleção de 7pb das sequências anteriores que gerou o frameshift, de forma semelhante à sequência de *H. parainfluenzae* T3T1. Portanto, o que teria ocorrido nestas sequências? Observamos, no alinhamento de nucleotídeos, que uma inserção de ponto de apenas uma guanina na posição 288 tirou as quatro sequências do frame (O0521, O1421, P2321 e HiE). É interessante notar que uma destas sequências, HiE, corresponde a um *H. influenzae*capsulado. Como as demais sequências correspondem a *H. influenzae* não tipáveis (que podem ser tanto um *H. influenzae* não capsulado, como um capsulado cuja cápsula é não tipável), nós concluimos que estas linhagens possuem uma cápsula não tipável e compartilham um ancestral comum capsulado com a linhagem capsulada HiE. Este ancestral capsulado sofreu o evento de inserção de ponto, passando esta característica às linhagens descendentes.

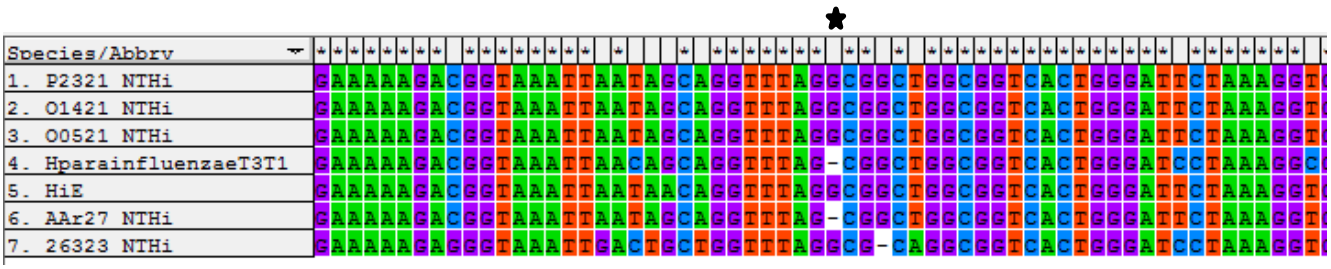


Figura 30: Alinhamento parcial das sequências P2321_NTHi, O1421_NTHi, O0521_NTHi e HiE (pseudogenes) com as sequências *H. parainfluenzae* T3T1 e 26323_NTHi (genes verdadeiros): eventos distintos leveram a pseudogeneização deste grupo. Notar a provável inserção de uma guanina (estrela preta) nas sequências HiE, P2321_NTHi, O1421_NTHi e O0521_NTHi.

Sob a ótica destas novas informações, alguns pontos previamente postulados na literatura devem ser considerados: em 1991, Kroll e colaboradores, ao identificarem e caracterizarem o gene da SodC nas espécies *H. influenzae* e *H. parainfluenzae*, afirmaram que o gene é bastante escasso na espécie *H. influenzae*(presente em apenas 17% das linhagens analisadas), o que foi confirmado neste trabalho, visto que na coleção particular de 45 linhagens de *H. influenzae*do Laboratório de Biotecnologia, apenas 3 linhagens possuem o gene, sendo que um é um pseudogene. Em contrapartida, o gene é bastante comum na espécie *H. parainfluenzae* estando presente em todas as linhagens analisadas. Estes mesmos autores afirmaram que todas as linhagens não tipáveis de *H. influenzae* possuíam a forma ativa da enzima, o que não parece ser verdadeiro, visto que algumas linhagens não tipáveis possuem um pseudogene e não um gene verdadeiro.

Além disso, estes autores afirmaram ainda que algumas linhagens capsuladas possuíam uma mutação de ponto na sequência do gene, a qual favoreceu a troca, na proteína, de uma Histidina na posição 82 por uma Arginina. Este resíduo seria fundamental para a ligação do íon cobre (Cu), um dos co-fatores da enzima juntamente com o íon zinco (Zn), de forma que esta substituição de aminoácido levou a inatividade desta enzima.Esta afirmação foi feita com base nasequência HiB, que também foi utilizada neste trabalho. Em primeiro lugar, Histidina e Arginina são aminoácidos de mesma classe e mesma polaridade (são aminoácidos básicos e, portanto, hidrofílicos). Esta troca de aminoácidos, a princípio, não seria drástica o suficiente para pertubar a ligação do co-fator. Em segundo lugar, avaliando-se a mesma sequência (Figura 31), foi observado que na posição 82 existe uma Tirosina. A Tirosina é um aminoácido polar neutro (ela não possui carga). Esta troca, de fato, poderia levar a uma ineficiência da ligação do co-fator, não só pela questão da afinidade entre as moléculas, mas também pelo fato de a Tirosina ser um aminoácido com

propensão a ser fosforilado. Portanto, não sabemos quais as repercussões da presença do grupamento fosfato na ligação do íon cobre.

>M84012_HiB

75 82

GLHGFHI**Y**ENPSCEPKEKDGKLIAGLAAGGHWDSKGAKQHGYPWQDDAHLGDLP

ALTVLHDGTATNPVLAPRLKKLDEVRGHSIMIHAGGDNHSDHPAPLGGGGPRMAC

GVIK*

Figura 31: sequência parcial da proteína SodC de HiB. Notar o aminoácido Tirosina destacado na posição 82.

Portanto, verifica-se que, apesar de *H. parainfluenzae* ter um gene *sodC* verdadeiro e este gene ser aparentemente bem sucedido neste espécie (presente em 100% das linhagens testadas em literatura (Kroll, 1991; McCrea, 2010;)), nas linhagens de *H. influenzae* que reberam o gene deste espécie, o gene aparentemente não foi bem sucedido, pois sofreu pseudogeneização na maioria das sequências pela presença de *indels.* Eventos desta natureza geralmente podem ser deletérios uma vez que podem alterar drasticamente a matriz de leitura do gene, modificando, consequentemente, a sequência e a conformação da proteína correspondente, com a consequente inativação da mesma. O que resta saber é: se o gene *sodC*confere um fenótipo adaptativo importante ao microorganismo, pois proporciona um mecanismo de evasão da resposta imune do hospedeiro, e a maioria das linhagens de *H. influenzae*, capsuladas ou não, não possuem o gene, de quais mecanismos alternativos estas linhagens lançaram mão para suprir esta carência?

Em 1988, Musser e colaboradores (Musser, 1988) fizeram um estudo da estrutura genética populacional da espécie *H. influenzae* e reconstruíram uma filogenia para a mesma, estabelecendo a seguinte hipótese de ancestralidade-descendência:a espécie é dividida em dois grupos, um ancestral, constituído por linhagens não tipáveis, e um grupo descendente (mais contemporâneo) constituído por linhagens capsuladas. O grupo descendente (linhagens capsuladas) dividem-se em dois grupos: Divisão I e Divisão II, sendo que a Divisão I corresponde ao grupo mais recente. O mesmo autor constatou posteriormente (Kroll, 1991) que o gene *sodC* era pouco prevalente no grupo ancestral (linhagens não tipáveis, especialmente as não capsuladas), e era mais escasso ainda no grupo contemporâneo. Neste último grupo, a Divisão I (a mais recente) não possui a versão do gene no seu genoma, enquanto que a Divisão II, possui o gene, porém este codifica uma proteína inativa. De acordo com este cenário, o gene *sodC* foi adquirido pelo grupo ancestral no entanto, a característica não foi bem sucedida na linhagem ancestral, tendendo a desaparecer nas linhagens recentes. Observa-se que o grupo mais contemporâneo (Divisão I) chega a não possuir a cópia do gene no seu genoma. Esta fato pode ser explicado da seguinte forma: a aquisição e, principalmente, a manutenção de mutações deletérias levaram a uma má adaptação do fenótipo em questão. Mutações deletérias costumam ser varridas do genoma, por seleção purificadora, para a manutenção da integridade da proteína, da enzima ou mesmo de toda uma rota metabólica da qual a enzima faz parte. Desta forma, a remoção de mutações deletérias e suas consequentes má adaptações tem uma implicação fundamental na manutenção da integridade e do *fitness* (aptidão) do próprio microorganismo em relação ao seu meio.

No caso da espécie *H. influenzae* uma má adaptação ocorreu, verificando-se o desaparecimento do gene na espécie contemporânea. É fato que o gene confere um fenótipo

importante à bactéria: permite o escape da fagocitose e, ao evadir da resposta imune, o parasita pode colonizar e/ou infectar o hospedeiro com sucesso. É fato também que este mesmo gene é bem sucedido nas outras espécies bacterianas nas quais ele ocorre, por exemplo: *H. parainfluenzae* (Kroll, 1991) *N. meningitis* (Dunn, 2003), *H. haemolyticus* (Fung, 2006*), Brucella abortus* (Steinmann, 1985), etc. Verificamos aqui, no entanto, que linhagens de *H. influenza* (incluindo as capsuladas, que são as mais virulentas, invasivas e responsáveis por causar meningite grave em crianças) aparentemente sobrevivem bem sem este fenótipo. Então surgem as seguintes perguntas: essas linhagens lançaram mão de outros mecanismos adaptativos para contornar esta situação? Quais seriam eles? Ou este fenótipo não é crucial para a sobrevivência, perpetuação e para o sucesso da bactéria?

Em 2010, McCrea e colaboradores (McCrea, 2010) constataram que, como o gene *sodC* era mais frequentemente encontrado em bactérias comensais do que em bactérias patogênicas, a enzima SodC funcionaria mais como um fator de colonização, permitindo o escape do reconhecimento por macrófagos da ororfaringe, do que como um fator de virulência, implicado na invasividade propriamente dita do microorganismo. Este fato é corroborado pelo fato de que o peróxido de hidrogênio, mesmo em pequenas quantidades, é capaz de paralisar o movimento ciliar do epitélio da mucosa respiratória, impedindo o "clearance" bacteriano (remoção da bactéria da ororfaringe) (Kroll, 1991). Sabe-se que a SodC, ao dismutar o ânion superóxido, gera peróxido de hidrogênio que é liberado no meio (se este não for convertido em água e oxigênio). Portanto, sob este prisma, a SodC seria, sim, um fator que contribui para a perpetuação do patógeno no seu sítio de colonização. Em contrapartida, a autora erra ao afirmar que este gene é mais encontrado em comensais, visto que o mesmo é amplamente distribuído entre bactérias altamente patogênicas, tais como: *E.*

coli, M. tuberculosis, Yersinia pestis, Shigella Flexineri, Salmonella entérica, etc. Desta forma, a história do gene *sodC*na espécie *H. influenzae* permanece bastante nebulosa.

5.3 Alinhamento das Sequências de Nucleotídeos e Reconstrução de uma Filogenia para o gene *sodC* de *N. meningitidis* juntamente com o gênero *Haemophilus*.

Com o intuito de tentar esclarecer qual espécie do gênero *Haemophilus* compartilhou o gene *sodC* com a espécie *N. meningitidis,* foi gerado um alinhamento contendo todas as sequências do gênero *Haemophilus*e da espécie *N. meningitidis,* e uma filogenia para o gene *sodC* nestas espécies foi reconstruída (Figura 32). Foram excluídas desta filogenia as sequências da espécie *H. ducrey* por se tratarem de sequências mais distantes, em termos de similaridade, e por possuírem características únicas não compartilhadas com as demais, como dito anteriormente. A árvore foi enraizada com uma sequência do gene *sodC*da espécie *Actinobacillus pleuropneumoniae.* O gênero *Actinobacillus*, juntamente com os gêneros *Pasteurella* e *Haemophilus*, compõem a família *Pasteurellaceae* (Tortora, 2000).. Portanto esta espécie foi considerada adequada para funcionar como grupo externo (Lemey, 2009).

Ao se observar esta filogenia, constatamos que as sequências de *N. meningitidis* compartilham um ancestral comum com as sequências de *H. influenzae* mais similares ao *H. haemolyticus.* Portanto, quanto à origem do gene *sodC* na espécie *N. meningitidis*, existem duas possibilidades igualmente possíveis: 1) o gene *sodC*na espécie *N. meningitidis* pode ter sido adquirido diretamente da espécie *H. haemolyticus,* 2) ou a espécie *H. haemolyticus* pode ter doado o gene para a espécie *H. influenzae*, que, por sua vez, doou o gene para a espécie *N. meningitidis.* Fato é que o gene de *N. meningitidis* compartilha em ancestral comum com o grupo de sequências mais aparentadas (ou mais similares) com o

gene de *H. haemolyticus*. Desta forma, o gene pode ter sido adquirido diretamente do *H. haemolyticus*, ouindiretamente através do *H. influenzae,* por evento de transferência lateral de genes como postulado na literatura (Kroll, 1998).

No entanto, todas as sequências pertencentes à espécie *N. meningitidis* agrupam num ramo distinto do ramo de *H. influenzae* e *H. haemolyticus*, verificando-se, com isto, que já houve suficiente acúmulo de variabilidade para tornar a sequência do gene de *N. meningitidis* diferente o bastante da sequência de *H. influenzae e H. haemolyticus* para que pudessem agrupar em ramos distintos da árvore. Isto pode ter acontecido por dois motivos: ou as sequências já estão divergindo há algum tempo, o que permitiu o acúmulo de variabilidade entre elas; ou o evento é muito recente na escala de tempo, porém pressões seletivas distintas muito fortes permitiram que as sequências se modificassem rapidamente.

Analisando-se o ramo da árvore que agrupa as sequências de *N. meningitidis,* constamos dois fatos interessantes: o primeiro deles é que as notas dos ramos internos são significativamente baixas, o que não sustenta a topologia deste ramos. Portanto, as relações evolutivas de parentesco entre estas sequências ainda não estão esclarecidas. No entanto, a nota do nó externo é bastante alta e sustenta a hipótese de divergência entre as sequências de *N. meningitidis* e *Haemophilus spp.*

O outro fato importante é que este ramo assemelha-se a um "pente". Este tipo de morfologia de ramo denomina-se Politomia e deve-se a uma Radiação Explosiva. Este fenômeno é particularmente comum entre vírus e ocorre quando um único ancestral dá origem a linhagens distintas quase que ao mesmo tempo (Lemey, 2009). Como estamos avaliando uma árvore de genes, estes dois fatos somados reforçam a hipótese de que o evento de transferência lateral foi muito precoce na escala de tempo. Principlamente em

casos de recombinação homóloga, a politomia reflete a incerteza da topologia precisa que melhor explica a relação evolutiva entre as sequências. (Lemey, 2009).

Ao se observar o alinhamento destas sequências (Figura 33), pode-se notar, a presença de um *indel* de 12pb (4 codons), portanto múltiplo de 3, não alterando, desta maneira, a matriz de leitura aberta do gene.

Procariotos não possuem introns, consequentemente não são capazes de fazer *splicing* alternativo (Tortora, 2000). O *splicing* alternativo constitui uma tática amplamente empregada por eucariotos para gerar novas formas de uma proteína à partir de um único gene, criando desta forma, novos alelos e, consequentemente, gerando variabilidade. Em procariotos, uma forma de se gerar variabilidade à nível da proteína é acrescentar ou remover da matriz de leitura do gene porções de sequência múltiplos de 3 (o que garante que o gene não ficará truncado. Não sendo múltiplo de 3, dar-se-ia, então origem a uma mutação do tipo *non sense,* e inevitavelmente o gene seria silenciado) (Lemey, 2009).

Observou-se ainda que vários sítios da sequência de *N. meningitidis* possuíam nucleotídeos distintos da sequência de *Haemophilus.* No entanto, para saber se o indel (inserção/deleção) foi o evento responsável pela diversificação das duas sequências, o mesmo foi removido de todas as sequências de *N. meningitidis*, um novo alinhamento foi feito e uma nova árvore filogenética foi gerada. Esta nova árvore não exibiu alteração de sua topologia em relação à árvore anterior. Portanto, observa-se que a diversificação entre as sequências de *N. meningitidis,* e *H. influenzae* e *H. haemolyticus* se deu muito mais pelo acúmulo de substituição de nucleotídeos ao longo da sequência ao longo do tempo, do que por efeito de um evento abrupto de inserção/deleção de uma única sequência de uma só vez.

Como as sequências do gene *sodC*em *N. meningitidis*são muito conservadas, quase não existindo variabilidade entre elas, supõe-se, então, que o evento de transferência lateral do gene foi muito recente no tempo. Desta forma, as sequências exibem alta clonalidade dentro de uma espécie, visto que não houve tempo o suficiente para que as mesmas pudessem acumular variabilidade.

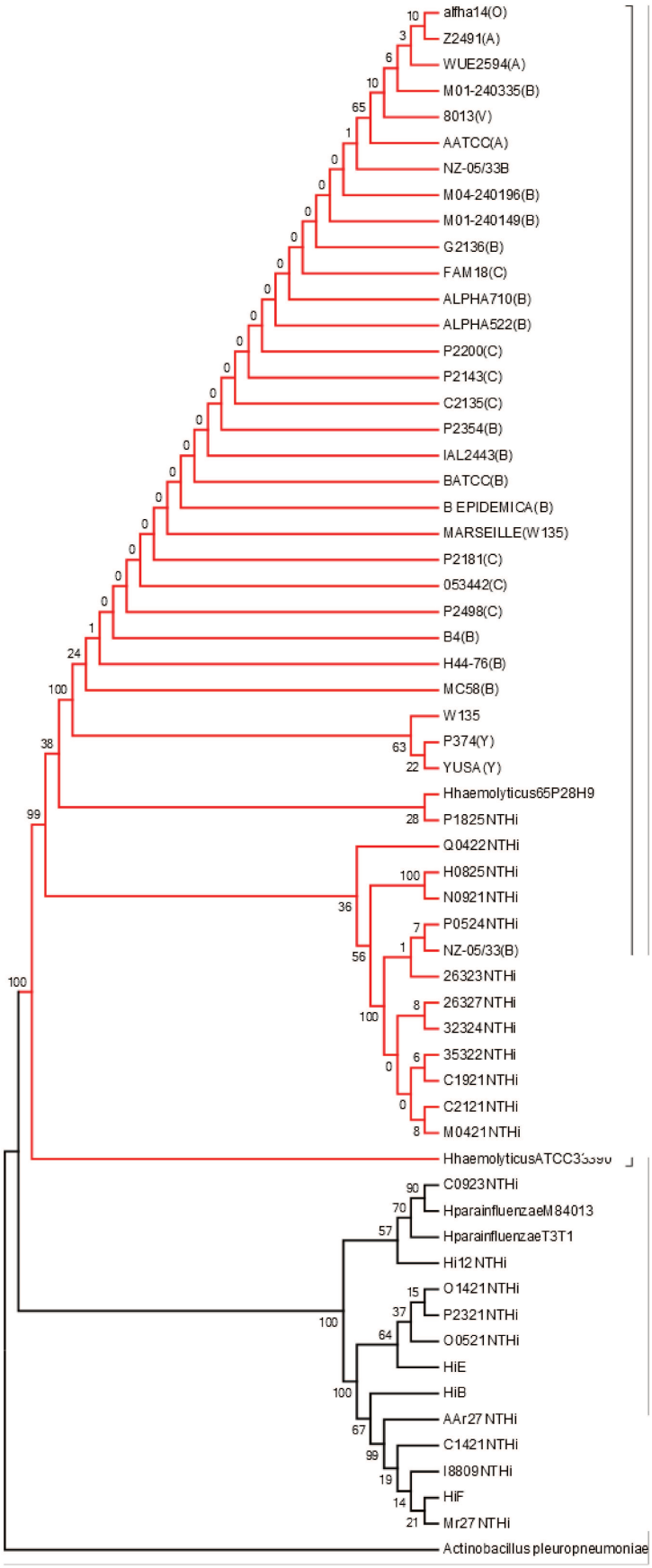

alfha14(O)
Z2491(A)
WUE2594(A)
M01-240335(B)
8013(V)
AATCC(A)
NZ-05/33B
M04-240196(B)
M01-240149(B)
G2136(B)
FAM18(C)
ALPHA710(B)
ALPHA522(B)
P2200(C)
P2143(C)
C2135(C)
P2354(B)
IAL2443(B)
BATCC(B)
B EPIDEMICA(B)
MARSEILLE(W135)
P2181(C)
053442(C)
P2498(C)
B4(B)
H44-76(B)
MC58(B)
W135
P374(Y)
YUSA(Y)
Hhaemolyticus65P28H9
P1825NTHi
Q0422NTHi
H0825NTHi
N0921NTHi
P0524NTHi
NZ-05/33(B)
26323NTHi
26327NTHi
32324NTHi
35322NTHi
C1921NTHi
C2121NTHi
M0421NTHi
HhaemolyticusATCC33390
C0923NTHi
HparainfluenzaeM84013
HparainfluenzaeT3T1
Hi12NTHi
O1421NTHi
P2321NTHi
O0521NTHi
HiE
HiB
AAr27NTHi
C1421NTHi
I8809NTHi
HiF
Mr27NTHi
Actinobacillus pleuropneumoniae

Figura 32: Árvore filogenética para as sequência do gênero *Haemophilus* e da espécie *N. meningitidis*.

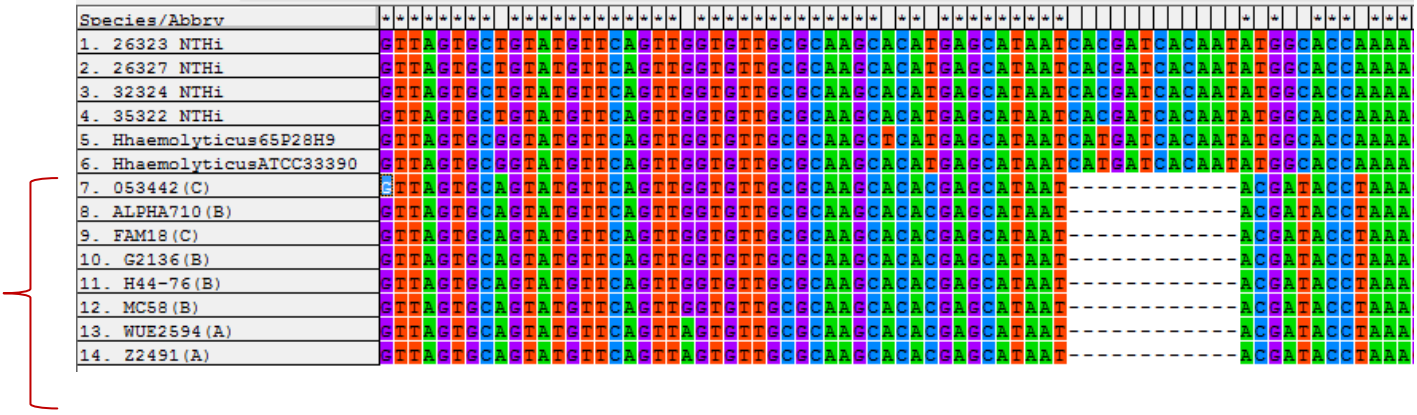


Figura 33: Alinhamento parcial de sequências representativas das espécies *N. meningitidis* (destacadas com um parênteses vermelho), *H. influenzae* e *H. haemolyticus*. Notar a provável deleção de 12pb nas sequências de *N. meningitidis*.

5.4 Teste de Seleção Positiva

A árvore utilizada como arquivo input para o programa CODEML do pacote PAML está ilustrada na Figura 34. Esta árvore contém sequências apenas do gênero *Haemophilus*, pois nas sequências da espécie *N. meningitidis* não foram encontradas substituições sinônimas, o que torna o cálculo do valor de ω, a base empregada pelo programa CODEML, uma impossibilidade matemática (Yang, 1997).

Supõem-se que esta espécie recebeu o gene por transferência lateral da espécie *H. haemolyticus*, direta ou indiretamente. Como o evento é muito recente na escala de tempo, ainda não houve tempo o suficiente para que uma diversificação molecular significativa tivesse ocorrido, o que é constatado observando-se a clonalidade destas sequências e a ausência de substituições sinônimas. Supomos, portanto, que a mesma força evolutiva que está atuando nas sequências de *H. haemolyticus*, também está moldando as sequências de

N. meningitidis. Para esclarecer o caráter das raras mutações encontradas nestas sequência, se vantajosas ou deletérias, faz-se necessário o emprego de modelagem proteica bem como a análise da dinâmica molecular.

Dentro de uma mesma população, os indivíduos estão sob a amesma pressão seletiva, o que tende a reduzir a variabilidade genética intrapopulacional e aumentar a variabilidade entre populações distintas. Portanto, diferentes populações podem experimentar diferentes pressões seletivas, aumentando a diferenciação entre estas populações e favorecendo a especiação. Seleção positiva favorece a modificação do estado ancestral, ao passo que seleção negativa mantém o fenótipo parental. Desta forma, a ação da seleção negativa tende a manter populações distintas mais similares. Como estamos tratando de genomas haplóides, não faremos menção à seleção do tipo balanceadora. (Lemey, 2009).

Portanto, como a espécie *H. influenzae* recebeu o gene *sodC* de duas fontes distintas, o objetivo desta etapa do trabalho foi avaliar se pressões seletivas distintas estão moldando padrões distintos de evolução molecular dentro da espécie*H. influenzae.* Dada a origem distinta do gene, poderiam forças seletivas distintas governar a evolução do mesmo gene dentro da mesma espécie?

Desta maneira, uma árvore com as sequências das espécies que compõem o gênero *Haemophilus* foi reconstruída. Desta árvore, foram removidas as sequências de *H. ducrey*por não haver uma amostragem significativa desta espécie. O tamanho diminuto da amostra pode levar a resultados equivocados. Também foi removido o ramo que continha as sequências correspondentes aos pseudogenes. Sabe-se que pseudogenes evoluem de uma forma diferente dos genes verdadeiros visto que não existe uma restrição funcional imposta pela proteína. Como pseudogenes não codificam uma proteína ou codificam uma proteína inativa, o acúmulo de substituições não-sinônimas ao longo da sequência chega a ser duas

vezes maior quando comparada a um gene verdadeiro, mesmo que este último esteja sob regime de seleçãopositiva. Portanto, supõem-se que a deriva genética molecular tem um efeito maior neste tipo de sequência.

A árvore foi enraizada pela sequência do gene *sodC* da espécie *Actinobacillus pleuropneumoniae.*. A árvore foi construída utilizando-se um método de distância (*Neighbor Joining* – Junção de Vizinhos), com modelo de substituição de nucleotídeos de Tamura Kumar, considerando-se as taxas de variação entre sítios uniformes e padrões homogêneos entre linhagens. O teste de confiabilidade amostral utilizado foi o de Reamostragem com Reposição (*Bootstrap*) com 1000 repetições e um limiar (*threshold*) de 50%. Este teste gerou notas altas o sufiente para se considerar a topologia desta árvore confiável.

Nesta árvore, os ramos testados estão destacados em vermelho e em preto. Estes ramos foram escolhidos porque compartilham ancestrais comuns com a espécie *H. haemolyticus,* e com a espécie *H. parainfluenzae*, respectivamente.

Os principais resultados desta análise estão dispostos nas tabelas14, 15 e 16.

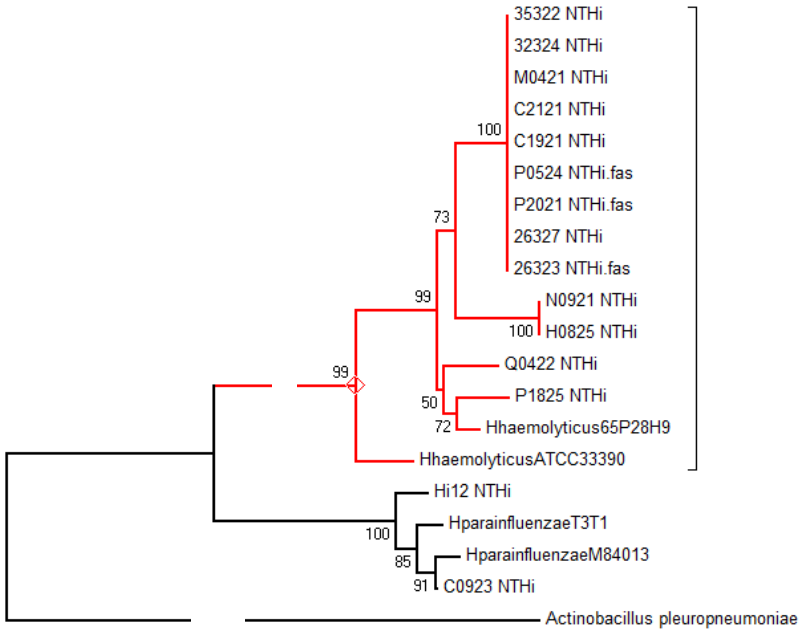


Figura 34: Árvore filogenética utiliza - mo arquivo de entrada para o programa CODEML.

Tabela 14: Resultados das análises de seleção pelos modelos de sítio do programa CODEML.

MODELO (sítios)	PARÂMETROS	*l*
M0	**ω = 0.061**	**-1665.490354**
M1a	**ω0 = 0.06099 e p0 =0.99999; ω1 = 1 e p1 = 0.00001**	**-1665.490383**
M2a	**ω0 = 0.061 e p0 = 1; ω1 = 1 e p1 = 0; ω2 = 22,32896 e p2 = 0**	**-1665.490354**

Tabela 15: Resultados das análises de seleção pelos modelos de ramos do programa CODEML.

RAMO	MODELO	PARÂMETRO	*l*	p-value
1	MAnulo	ω0 = 0.05071 e p0 = 0.88384 ω1 = 1 e p1 = 0.00269 ω2 = 1 e p2 = 0.11313	-1662.489656	
	MAalternativo	ω0 = 0.05220 e p0 = 0.96717 ω1 = 1 e p1 = 0 ω2 = 478,13502 e p2 = 0.03283	-1660.199949	0.03247
6	MA nulo	ω0 = 0.05491 e p0 = 0.92392 ω1 = 1 e p1 = 0.00140 ω2 =1 e p2= 0.07457	-1664.880472	
	MAalternativo	ω0 = 0.5388 e p0= 0.96401 ω1 = 1 e p1 = 0.0137 ω2 = 3.91237 e p2 = 0.03458	-1664.675002	

Tabela 16: Resíduos de aminoácidos com probabilidade de estar sob seleção positiva. A significância da estatística foi dada pelo teste de Bayes Empirical Bayes (BEB).

RAMO	RESÍDUO DE AA SOB SELEÇÃO POSITIVA	BEB	SIGNFICATIVO
1	3 (Alanina) (A)	0.609	NÃO
	6 (Serina) (S)	0.598	NÃO
	11 (Glutamina) (Q)	0.633	NÃO
	16 (Histidina) (H)	0.698	NÃO
	18 (Lisina) (K)	0.656	NÃO
	57 (Serina) (S)	0.989	SIM
6	11 (Glutamina) (Q)	0.652	NÃO
	16 (Histidina) (H)	0.635	NÃO
	18 (Lisina) (K)	0.680	NÃO

O modelo M0 (*one ratio*) calcula um único valor de ω para toda a árvore. Neste caso o valor de ω = 0.061, portanto seleção negativa atua fortemente ao longo das linhagens desta árvore bem como ao longo dos sítios das sequências. Este resultado corrobora com a literatura no sentido de que, dada a restrição ou imposição funcional que existe sobre moléculas de proteínas, e, consequentemente sobre a sequência de DNA qua a codifica, a

seleção negativa atua sobre a molécula durante a maior parte do tempo evolutivo e ao longo da maioria dos sítios para manter íntegra a função da proteína ou enzima. Como a seleção negativa tem este caráter: de atuar ao longo da molécula durante um período de tempo evolutivo maior e ao longo da maioria dos sítios, ela pode apagar ou diluir o efeito da seleção positiva, já que esta última é mais pontual, atuando em poucos resíduos específicos, porém cruciais, e durante um período de tempo menor (Yang, 1994).

Desta forma foi testado um par de modelos de sítios: M1a X M2a. Estes modelos também atribuem notas para toda a árvores, porém eles permitem que os sítios variem de caráter. O modelo M1a (neutro e, portanto, menos complexo) assume dois tipos de sítios de caráteres distintos: sítios p0 (conservados), nos quais $\omega 0 < 1$, e sítios p1, nos quais $\omega 0 = 1$. Portanto os valores de ω assumem o caráter de uma distribuição de probabilidades, na qual ω será sempre 1 numa dada proporção de sítios p1, e ω terá um valor entre 0 e 1 numa dada proporção de sítios p0. O modelo M2a (modelo de seleção positiva e, portanto, mais complexo) incorpora uma terceira classe de sítios p2, nos quais $\omega 2 > 1$. Ao se aplicar o Teste da Razão de Máxima Verossimilhança (LRT), verificamos que a diferença da nota entre os dois modelos não é significativa, portanto o modelo mais complexo (o que aceita seleção positiva) não é melhor que o modelo mais simples (modelo de neutralidade) para explicar os dados. Desta forma, a hipótese alternativa foi rejeitada. Observamos, no modelo M2a que $\omega 2 = 22.33$, porém numa proporção de sítios p2 =0, ou seja, na árvore, como um todo, não existem sítios positivamente selecionados. Portanto, tem-se que 99% dos sítios estão sob forte regima de seleção negativa com $\omega = 0.06099$. Estes resultados estão dispostos na tabela 8.

Diante disto, na tentativa de identificar algum resíduo de aminoácido que estivesse sob seleção positiva, realizamos testes com os modelos de ramos, que permitem que os

valores de ω variem dentre as linhagens da árvore (tabela 9). O primeiro teste foi feito para o ramo 1. Nele observamos que a nota do modelo alternativo (modelo de seleção positiva) é significativamente maior que a nota do modelo nulo. O teste de *LRT* (comparação entre os dois modelos) rejeitou a hipótese nula, de forma que o modelo alternativo (seleção positiva) explica melhor os dados. O teste do x^2 modela a distribuição de probabilidades dos valores de modelos aninhados e foi usado para atestar a significância estatística do teste de *LRT* (p-value = 0.03247).

Desta forma, temos que uma proporção de 3% de sítios p2 tem um valor de ω = 478,13. Este resultado corroboraa literatura, visto que a seleção adaptiva ou diversificadora atua sobre poucos resíduos de aminoácidos, porém cruciais. O teste de Bayes Empirical Bayes (BEB) apontou que o único aminoácido com forte probabilidade de estar sob seleção positiva é o da posição 57 (BEB = 98%) que neste ramo corresponde a uma serina (tabela 10). Nesta mesma posição, nas sequências que constituem o outro ramo da árvore existe uma alanina. Portanto, constatamos que a substituição de uma alanina por uma serina na posição 57 foi vantajosa. Alanina é um amominoácido hidrofóbico, enquanto que serina é um aminoácido polar neutro, porém a serina é um aminoácido que frequentemente é fosforilado. O grupo fosfato é carregado negativamente, e a fosforilação ou desfosforilação de um resíduo de serina tem fortes implicações nas vias de sinalização intracelular.

Em relação ao ramo 6, observamos um ω2 = 4 em 3% da molécula. Porém, o modelo alternativo não foi melhor que o modelo nulo para explicar os dados, portanto, numa análise parcial deste ramo, temos que em 92% da sequência ω = 0.05.

Desta forma, entre as sequências que compartilham um ancestral comum com o *H. haemolyticus* houve variação e seleção adaptadora atuou sobre esta variação mantendo-a. Já as sequências que compartilham um ancestral comum com o *H. parainfluenzae* estão sob

um regime de forte seleção purificadora na maior parte da sequência. Esta força tende a varrer novidades evolutivas da sequência no intuito de preservar a função da proteína.

Para se realizar a análise da espécie *N. meningitidis* separadamente um novo alinhamento foi gerado. No entanto, notou-se a ausência de substituições sinônimas ao longo do alinhamento, como explicado previamente no item 4.1. Como toda a análise gerada pelos modelos implementados pelo programa CODEML baseiam-se no cálculo do valor da razão $\omega = dN/dS$, a análise tornou-se inviável dado que o numerador do equação é zero (dS = 0), o que gera uma impossibilidade matemática. No entanto, não podemos afastar nem a hipótese nula, nem a hipótese alternativa já que o teste estatístico não pode ser feito. Diante disto, temos três alternativas: a primeira seria procurar mais sequências divergentes na tentativa de introduzir variabilidade ao alinhamento, o que foi feito conforme explicado no item 4.1. No entanto, estas novas sequências, além de não introduzirem variabilidade ao alinhamento, diminuíram o poder do mesmo, visto que as sequências eram menores. Como explicado anteriormente, quanto menor as sequências, menor o poder de comparação entre elas. A segunda alternativa seria analisar os sítios ondehouve troca de aminoácidos, modelar a estrutura molecular da SodC e identificar possíveis consequências na função da enzima devido a estas trocas de aminoácidos. A terceira alternativa seria utilizar um programa que não utilize o dN e o dS como princípios da identificação de seleção.

Notamos o seguinte: ambos os testes de seleção positiva ao longo de linhagens (*branch models*) e o teste de seleção positiva nos sítios (*site models*) de aminoácidos são altamente conservativos. Quando testamos linhagens sob seleção positiva, assumimos que a razão ω é idêntica ao longo dos sítios. Seleção positiva, então, somente é detectada ao longo de uma linhagem somente se a média da razão ωsobre todos os sítios é

significativamente maior que 1. Dado que muitos sítios numa proteína, ou a maioria deles, estão sob a ação da seleção purificadora, e, portanto, o valor subjacente de *w* será muito próximo de zero, este teste é, desta forma, muito conservativo para testar seleção positiva (Yang, 1994). De modo semelhante, o teste da razão de máxima verossimilhança para sítios positivamente selecionados baseia-se na suposição de que a razão w é idêntica em todas as linhagens numa árvore. Seleção positiva somente é detectada num sítio se a média da razão w subjacente sobre todas as linhagens é significativamente maior que 1. Esta suposição parece pouco realista dado que, para a amioria dos genes, seleção positiva afeta somente algumas poucas linhagens (Yang, 1997).

Além disso, os modelos aqui discutidos pressupõema mesma razão ω para qualquer troca de aminoácido. Portanto, assume-se que, para um sítio positivamente selecionado, a mudança de um aminoácido para qualquer outro é considerada vantajosa. Como sabemos, tal suposição é irreal para qualquer proteína (Yang, 2000). Soma-se a isto, o fato de que os testes aqui utilizados somente detectam seleção diversificadora se a taxa de substituição não sinônima é maior que a taxa de substiuição sinônima. No entanto, sabe-se que uma única mudança de aminoácido pode ser adaptativa mesmo que $dN/dS < 1$ (McClellan, 2009). Portanto, enquanto que o programa pareça muito conservativo para detectar seleção positiva em todos os casos, ele é robusto o suficiente para detectar seleção positiva em casos como o gene da lisozima e do gene *vif* do vírus HIV-1. Nestes casos, seleção positiva foi detectada mesmo que a média da razão $w < 1$ para todos os sítios.

Um passo importante para aumentar o poder do teste da razão de máxima verossimilhança é desenvolver modelos que permitam que a razão w varie ao longo de linhagens a ao longo de sítios. Um modelo computacional com estes requisitos parece ser viável, embora muito mais «computação» e tempo sejam necessários. Por um lado, é

importante focar num curto período de tempo e em poucos resíduos de aminoácidos de forma que o sinal da seleção adaptativa não seja subestimado pelo efeito da seleção purificadora, a qual atua num período de tempo maior e sobre a maioria dos sítios numa proteína. Por outro lado, um período muito curto de tempo e poucos resíduos de aminoácidos talvez não ofereçãm espaço suficiente para que o sinal da seleção positiva seja detectado por meio de testes estatísticos (Yang, 2000).

5.5 Modelagem Proteica Computacional por Homologia

Partes dentro de um mesmo gene são mais conservadas do que outras, principalmente aquelas que codificam sítios catalíticos, "core" protéicos, alfa hélices, folhas beta e resíduos de aminoácidos responsáveis pela formação de pontes intramoleculares. Entre proteínas homólogas, porém distantemente relacionadas, a similaridade entre suas sequências primárias é perdida mais rapidamente do que a semelhança entre suas estruturas secundária e terciária, as quais são condicionantes para o bom funcionamento de uma determinada enzima (Lemey, 2009)

Esta questão é de fundamental importância para a compreensão de como a seleção natural atua a nível molecular e de como ela molda a diversificação entre os organismos. A seleção negativa ou purificadora atua sobre a maior parte de uma sequência durante a maior parte de sua história evolutiva. Isto se dá porque a imposição ou restrição funcional proteica sobre a molécula é predominante. Isto quer dizer que um organismo é um "emaranhado" de rotas metabólicas muito bem ordenadas, com inúmeras proteínas sinalizando e interagindo umas com as outras ao mesmo tempo para manter a integridade bem como a homeostasia daquele organismo. Uma única mutação, levando a uma mudança drástica de aminoácido (classes distintas, polaridades diferentes, etc) pode ter uma repercussão desastroza sobre

uma única proteína (alterando o seu *folding* e inativando o seu sítio catalítico) bem como sobre toda uma rota metabólica da qual aquela proteína tenha uma participação crucial. Desta forma, seleção negativa parece ser a força predominante. Quando uma mudança de amoniácido é vantajosa para uma proteína porque irá melhorar a sua performance, então esta modificação é aceita na sequência da molécula e aforça responsável por introduzir esta mudança adaptativa é a seleção positiva. A seleção positiva tem um caráter mais pontual. Atua sobre poucos aminoácidos por vez, porém em resíduos chaves para o melhor funcionamento da proteína em relação ao seu meio. A seleção positiva tem um caráter mais preponderante em casos particulares em moléculas que precisam de muita variabilidade para o seu bom funcionamento, como moléculas do MHC, proteínas da suérfície de patógenos, anticorpos, etc (Yang, 1994)

Portanto, a substituição de uma purina (A,G) por outra purina e de uma pirimidina (C,T) por outra pirimidina ocorre mais facilmente por razões químicas e estéricas. Tais mutações são denominadas de transições. Uma transversão ocorre quando uma purina é substituída por uma pirimidina. O inverso é menos comum pelos mesmos motivos acima. No que tange à modificação do aminoácido à nível da proteína, transversões costumam ter um impacto maior do que transições, porque resultam em alterações mais drásticas das propriedades químicas do aminoácido codificado. Transversões tem maior chance de ocorrer do que transições, no entanto, o código genético evoluiu de tal forma que transições (menos disruptivas, portanto com menor restrição química) são mais comuns do que transversões. (Lemey, 2009).

A mutação de ponto num códon, em particular, geralmente troca o aminoácido para aquele com propriedades químicas semelhantes, de forma que a estrutura secundária e

terciária da proteína codificada não seja alterada drasticamente. Portanto, organismos vivos toleram apenas tipos limitados de mutações nas regiões codificantes do seu genoma. Vale ressaltar que as mutações ocorrem ao acaso, sendo as forças evolutivas as responsáveis por limitar o tipo de mutação e moldar as sequências da forma como as enxergamos. (Lemey, 2009)

Com base no proteoma da linhagem MC58 da bactéria *N. meningitidis*, sorogrupo B (Wilks, 1998; Tettelin, 2000), a SodC constitui um homodímero e possui 186 AA, sendo que os 22 primeiros são altamente hidrofóbicos e correspondem a um peptídio sinal (Figura 35). A enzima, que faz parte da família das metaloenzimas, exibe atividade de oxido-redutase, desprotonando o íon cobre (metal com função catalítica) para oxidar ânions superóxidos, segundo a seguinte reação: 2 superoxido + 2 H^+= O_2+ H_2O_2.

```
        10         20         30         40         50         60
MNMKTLLALA VSAVCSVGVA QAHEHNTIPK GASIEVKVQQ LDPVNGNKDV GTVTITESNY

        70         80         90        100        110        120
GLVFTPDLQG LSEGLHGFHI HENPSCEPKE KEGKLTAGLG AGGHWDPKGA KQHGYPWQDD

       130        140        150        160        170        180
AHLGDLPALT VLHDGTATNP VLAPRLKHLD DVRGHSIMIH TGGDNHSDHP APLGGGGPRM

ACGVIK
```

Figura 35: Sequência primária da SodC de *N. meningitidis.*

O íon Zn cumpre papel estrutural, sendo ligado pelos aminoácidos 104, 113, 122 e 125. Os resíduos 86 e 182 formam pontes dissulfeto intramoleculares. Os resíduos 79, 81, 104 e 160 ligam o íon cobre e, portanto exibem atividade catalítica. A estrutura

tridimensional das SodsC procariótica e eucarótica sugere que estas enzimas evoluíram por caminhos distintos (Kroll, 1991).

No intuito de verificar se a substituição de uma alanina por uma serina na posição 57 do alinhamento de nucleotídeos, conforme exposto no item anterior, confere alguma vantagem `a enzima SodC, foi feita uma modelagem compotacional com base em Homologia. O arquivo de entrada utilizado foi o 2AQN do banco de dados PDB, que corresponde à enzima SodC de *N. meningitidis*, sorogrupo B (Figura 36).

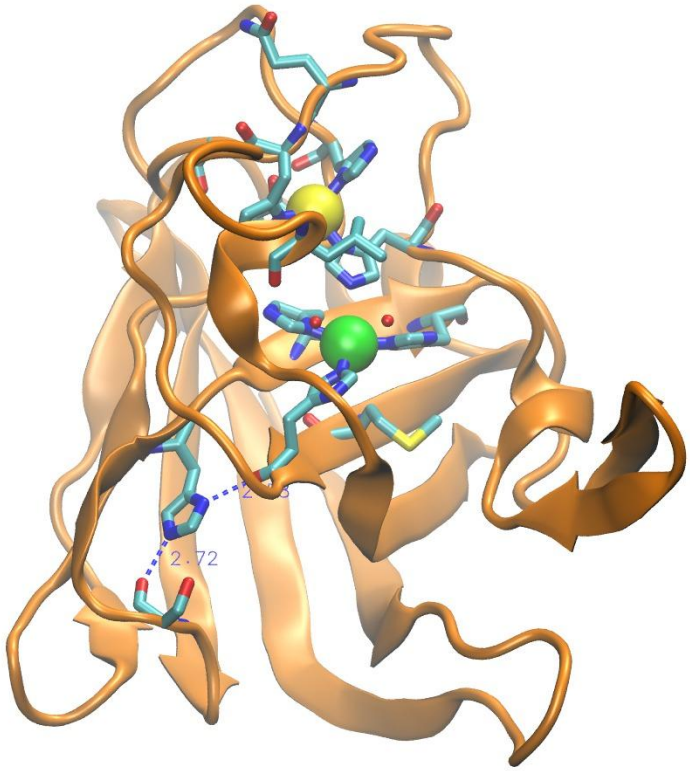


Figura 36: Modelagem computacional das repercussões da mutação alanina→serina sobre a dinâmica estrutura da enzima SodC.

Foi observado que a presença de uma serina na posição 72 da estrutura primária da proteína, que corresponde à posição 57 do alinhamento de nucleotídeos, a 2.72A da Histidina 76, forma uma rede de ligações de hidrogênio bem estruturada (em azul pontilhado), via Histidina 76, estabilizando-a. Esta, por sua vez, a 2.58A da Histidina 160, polariza esta Histidina, um dos aminoácidos que compõem o sítio catalítico desta enzima e também um dos resíduos que liga o íon cobre, o qual, na condição de íon catalítico, é reponsável pela troca de elétrons durante a dismutação do ânion superóxido A forma da enzima que contém a alanina nesta posição perde essa rede bem estruturada, alterando o potencial eletroquímico do sistema, e, consequentemente, alterando o potencial redox do sítio catalíco.

Portanto, verifica-se que a substituição de uma alanina por uma serina, nas vizinhanças do sítio ativo, favorece, indiretamente a atividade da enzima, estabilizando a estrututa do seu sítio catalítico e melhorando o seu potencial redox..

6. Conclusões

Acerca da dinâmica do gene *sodC* nas espécies consideradas neste trabalho, conclui-se que:

- ✓ as sequências pertecentes à espécie *N. meningitidis* são altamente clonais, chamando a atenção a ausência de substituições sinônimas e algumas poucas substituições não sinônimas. Isto nos faz supor que estas sequências estão divergindo há muito pouco tempo, portanto o evento de transferência lateral de genes é muito recente na escala de tempo.Uma filogenia para estas sequências foi reconstruída e a nota atribuída aos ramos não suporta a sua topologia, especialmente porque as poucas mutações existentes não formam um padrão de distribuição ao longo de todas as sequências. Este fato, portanto, suporta a hipótese de que o evento de transferência lateral foi realmente muito recente na escala do tempo. Não foi possível aplicar o teste de seleção positiva baseado na estatística $\omega = dN/dS$ para esta espécie. Como o teste não pode ser aplicado, não podemos suportar a hipótese nula (seleção positiva não está atuando) bem como não podemos rejeitar a hipótese alternativa (seleção positiva está atuando).

- ✓ Em relação ao gênero *Haemophilus,* observou-se que uma parte das linhagens de *H. influenzae* NTHi compartilha um ancestral comum com as linhagens de *H. haemolyticus.* Como eventos de transferência lateral de genes são muito comuns neste gênero acredita-se que a espécie *H. haemolyticus* possa ter transferido o gene para a espécie *H. influenzae* ou vice-versa. Observou-se também qua a outra parte das linhagens de *H. influenzae* NTHi juntamente com as linhagens capsuladas de *H.*

influenzae compartilham um ancestral comum com a espécie *H. parainfluenzae*. Novamente, nós acreditamos que estas espécies possam ter trocado o gene da SodC entre si por transferência lateral e recombinação homóloga. Foi constatado ainda que, neste ramo, as sequências de *H. influenzae* tornaram-se pseudogenes devido à presença de um *indel* de 7pb que, por não ser múltiplo de 3, tirou a matriz de leitura aberta do frame acarretando no surgimento de vários stop códons prematuros ao longo da sequência. É digno de nota que, os testes de confiabilidade amostral geraram notas altas o suficiente para corroborar esta hipótese. Portanto, nós constatamos que a espécie *H. influenzae* recebeu o gene *sodC* de duas fontes distintas: uma do *H. haemolyticus* e outra do *H. parainfluenzae.*

- ✓ As sequências de *N. meningitidis* compartilham um ancestral comum com a espécie *H. haemolyticus*e e com algumas linhagens NTHi de *H. influenzae*. Dado que existe uma alta probabilidade da espécie *H. influenzae* ter recebido o gene da espécie *H. haemolyticus,* nós supomos que a espécie *N. meningitidis* pode ter recebido o gene *sodC* diretamente da espécie *H. haemolyticus* ou indiretamente, através da espécie *H. influenzae*. Uma outra constatação também nos parece bastante razoável.

- ✓ As sequências de *N. meningitidis*, agrupam num ramo distinto dos *Haemophilus*. Portanto, já se observa algum tipo de diversificação molecular entre ambas.

- ✓ Um teste de seleção positiva constatou que seleção negativa é a força predominante entre as sequências. Portanto, existe uma força de caráter determinístico que mantém a estrutura da enzima sem muitas novidades evolutivas em termos de

substituições de aminoácidos. Se novidades adaptativas, na forma de mutações, estão sendo "varridas" da sequência significa que uma restrição ou imposição funcional muito forte requer que a enzima mantenha a sua estrutura inalterada de forma a não pertubar o seu funcionamento.

- ✓ De forma interessante, foi detectada seleção positiva atuando apenas no ramo que compartilha um ancestral comum com a espécie *H. haemolyticus,* especificamente numa serina que ocupa a posição 57 do alinhamento (BEB =98%). Foi constatado que o valor de ω atribuído a este ramo foi bastante alto (ω = 478.0) se comparado com dados da literatura, o que sugere que este aminoácido pode ter uma atribuição crucial para o funcionamento da enzima como um todo.

- ✓ Desta forma constatamos que padrões distintos de evolução molecular estão ocorrendo no gene *sodC* da espécie *H. influenzae:* as linhagens que se tornaram pseudogenes, por não haver nenhum tipo de restrição funcional (visto que a enzima é inativa (Kroll, 1991)), estão sob uma ação maior da deriva genética. As demais linhagens sofrem preponderantemente a ação da seleção puriricadora, sendo que, neste grupo, a substituição de uma alanina por uma serina foi extremamente vantajosa.

- ✓ Em relação à espécie *N. meningitidis,* sabe-se que ela agrupa no mesmo ramo onde está ocorrendo seleção positiva, no entanto as suas sequências não puderam ser analisadas devido a uma impossibilidade matemática gerada pelo programa. SE a hipótese de TLG estiver correta, a mesma força evolutiva que opera sobre estas sequências de *H. influenzae* e *H. haemolyticus*, também opera sobre as linhagens de

N. meningitidis, principalmente porque o evento é recente na escala de tempo. No entanto, faz-se necessário a utilização de programas de modelagem proteica computacional para identificar as possíveis substituições de aminoácidos, suas implicações estruturais e funcionais, bem como suas repercussões na filogenia destes grupos, apontando quais os resíduos de aminoácidos cruciais para o evento de diversificação molecular.

- ✓ A modelagem proteica evidenciou que a troca do AA alanina por uma serina na posição 72 da estrutura primária de enzima foi vantajosa por, indiretamente, melhorar o potencial redox do sítio catalítico.

7. Referências Bibliográficas

Aguileta G, Refrégier G, Yockteng R, Fournier E, Giraud T. 2009. Rapidly evolving genes in pathogens: methods for detecting positive selection and examples among fungi, bacteria, viruses and protists. Infect Genet Evol. 9 (4): 656-70.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.1990. Basic local alignment search tool.J Mol Biol. Oct 5;215(3):403-10.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. Sep 1;25(17):3389-402. Review.

Antignac A., Alonso J.M., Taha M.K. 2000. Le point sur la résistance de *Neisseria meningitidis* aux antibiotiques, d'après les souches étudiées au Centre National deRéférence des Méningocoques en 1998. Antibiotiques 2: 241-245.

Aspholm M., Aas F.E., Harrison O.B., Quinn D., Vik A., Viburiene R., Tønjum T., Moir J., Maiden M.C., Koomey M. 2010. Structural alterations in a component of cytochrome c oxidase and molecular evolution of pathogenic *Neisseria* in humans.PLoS Pathog. 2010, 19;6 (8): e1001055**.**

Ausubel F.M, Brent R, Kingston, R.E. 1988. Current Protocols in Molecular Biology. Greene Publishing Associates, Brookling.

Baethgen LF, Moraes C, Weidlich L, Rios S, Kmetzsch CI, Silva MS, Rossetti ML, Zaha A. 2003 Direct-test PCR for detection of meningococcal DNA and its serogroup characterization: standardization and adaptation for use in a public health laboratory.
J Med Microbiol. Sep;52(Pt 9):793-9.

Balding D. J., Bishop M., Cannings C. 2007. Handbook of Statistical Genetics. 3rd ed, cap.12. Wiley.

Balakrishnan Priya, Reddi K Sivaprasanth, Vincent Dhivya Jensi, Lakshmanan Uma, Gopalakrishnan Subramanian, Dharmar Prabaharan. 2010. Characterization of manganese superoxide dismutase from a marine cyanobacterium *Leptolyngbya valderiana.* Saline Systems. 6: 6.

Banister J. V. and Parker M. W. 1985. The presence of a copper-zinc superoxide dismutase EC1.15.1.1 in the bacterium *Photobacterium leiognathi*; a likely case of gene transfer from eukaryotes to prokaryotes. Proc. Natl. Acad. Sci. USA. 82, 149-152.

Beauchamp C. O. and Fridovich I. 1971. Superoxide dismutase: improved assay and an assay applicable to acrylamide gels. Anal Biochem. 72, 276-287.

Biswas GD, Burnstein KL, Sparling PF.1986. Linearization of donor DNA during plasmid transformation in *Neisseria gonorrhoeae*. J. Bacteriol. 168: 756-61.

Burman WJ, Martin WJ 1986. Oxidant-mediated ciliary dysfunction. Possible role in airway disease. Chest. Mar;89(3):410-3.

Chillemi G 2012. Carbon monoxide binding to the heme group at the dimeric interface modulates structure and copper accessibility in the Cu,Zn superoxide dismutase from Haemophilus ducreyi: in silico and in vitro evidences. Biomol Struct Dyn,

Chen I; Dubnau D .2004. DNA uptake during bacterial transformation.Nat Rev Microbiol. 2(3):241-9. Review.

Chen SL, Hung CS, Xu J, Reigstad CS, Magrini V, Sabo A, Blasiar D, Bieri T, Meyer RR, Ozersky P, Armstrong JR, Fulton RS, Latreille JP, Spieth J, Hooton TM, Mardis ER, Hultgren SJ, Gordon JI. 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli:* a comparative genomics approach.Proc Natl Acad Sci U S A. 11;103 (15): 5977-82.

Claverys, J. P; Havarstein, L. S. 2002. Extracellular-peptide control of competence for genetic transformation in Streptococcus pneumoniae. Front. Biosci. 7, d1798–d1814.

Claverys, J. P; Martin, B. 2003 Bacterial 'competence' genes: signatures of active transformation, or only remnants? Trends Microbiol. 11, 161–165.

Dolan-Livengood J.M., Miller Y.K., Martin L.E., Urwin R., Stephens D.S.2003. Genetic basis for nongroupable *Neisseria meningitidis*. J Infect Dis. 15;187(10):1616-28.

Doulet N. D. E., Laran-Chich M.P., Niedergang F., Nassif X., Couraud P.O., Bourdoulous S. 2006. *Neisseria meningitidis* infection of human endothelial cells interferes with leukocyte transmigration by preventing the formation of endothelial docking structures. J Cell Biol. 22;173(4):627-37.

Dubnau D. 1999. DNA uptake in bacteria.Annu Rev Microbiol. 53:217-44. Review.

Dubnau D, Cirigliano C. 1972. Fate of transforming DNA following uptake by competent *Bacillus subtilis.* III. Formation and properties of products isolated from transformed cells which are derived entirely from donor DNA. J. Mol. Biol. 64: 9-29.

Dubnau D, Cirigliano C. 1972. Fate of transforming DNA following uptake by competent *Bacillus subtilis.* IV. The endwise attachment and uptake of transforming DNA. J. Mol. Biol. 64:31-46.

Dunn K. L., Farrant J. L., Langford P. R., Kroll J. S. 2003. Bacterial [Cu,Zn]-cofactored superoxide dismutase protects opsonized, encapsulated *Neisseria meningitidis* from phagocytosis by human monocytes/macrophages. Infect Immun. 71(3):1604-7.

Ewing, B.; P. Green. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research, 8(3):186-194.

Facius D, Fussenegger M, Meyer TF. 1996. Sequential action of factors involved in natural competence for transformation of *Neisseria gonorrhoeae*. FEMS Microbiol. Lett. 137:159–64

Facius D, Meyer TF. 1993. A novel determinant (comA) essential for natural transformation competence in *Neisseria gonorrhoeae*. Mol. Microbiol. 10:699–712.

Falla, TJ. 1994. PCR for capsular typing of *Haemophilus influenza.* J Clin Microbiol. 32:2382-6.

Findlay W. A. and Redfield R. J. 2009. Coevolution of DNA uptake sequences and bacterial proteomes. Genome Biol Evol. 5;1:45-55.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM. 1995 Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science. Jul 28;269(5223):496-512.

Flohé L. and Otting F. 1984. Superoxide dismutase assays. Methods Enzymol. 105: 93-104.

Foxwell, AR. 1998. Nontypeable *Haemophilus influenzae*: pathogenesis and prevention. Microbiol Mol Biol Rev. 62:294-308.

Freeman, D; Heron, SC. 2007. Evolutionary Analysis. 4 ed. Pearson Education LTD, London.

Fridovich I. 1986. Superoxide dismutases. Adv. Enzymol. 58: 61-97.

Fung WW,O'Dwyer CA, Sinha S, Brauer AL, Murphy TF, Kroll JS, Langford PR. 2006. Presence of copper- and zinc-containing superoxide dismutase in commensal *Haemophilus haemolyticus* isolates can be used as a marker to discriminate them from nontypeable *H. influenzae* isolates. J Clin Microbiol. Nov;44(11):4222-6.

Grace S. C. 1990. Phylogenetic distribution of superoxide dismutase supports and endosymbiotic origin for chloroplasts and mitocondria. Life Sci. 47: 1875-1886.

Haijema BJ, van Sinderen D, Winterling K, Kooistra J, Venema G. 1996. Regulated expression of the dinR and recA genes during competence development and SOS induction in *Bacillus subtilis*. Mol. Microbiol. 22:75–85.

Hamoen, L. W., Venema, G; Kuipers, O. P.2003. Controlling competence in Bacillus subtilis: shared use of regulators. Microbiology 149, 9–17.

Hassan H.M. 1989. Microbial Superoxide dismutases. Adv. Genet. 26: 65-97.

Hitchcock P. 1989. Unified nomenclature for pathogenic *Neisseria* species. Clinical Microbiology Reviews. 2, Suppl.: S64-S65.

Hill D. J., Griffiths N. J., Borodina E. 2010. Cellular and molecular biology of *Neisseria meningitidis* colonization and invasive disease. Clinical Science. 118: 547-564.

Kilian, M. 1976. A taxonomic study of the genus *Haemophilus,* with the proposal of a new species. J Gen Microbiol, 93: 9-62.

Krieger E, Joo K, Lee J, Lee J, Raman S, Thompson J, Tyka M, Baker D, Karplus K. 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77 Suppl 9,114-122.

Kroll J. S., Langford P.R., Loynds B.M. 1991. Copper-zinc superoxide dismutase of *Haemophilus influenzae* and *H. parainfluenzae*.J Bacteriol. 173(23):7449-57.

Kroll J.S., Langford P.R., Wilks K.E., Keil A.D. 1995. Bacterial [Cu,Zn]-superoxide dismutase: phylogenetically distinct from the eukaryotic enzyme, and not so rare after all!Microbiology.141 (Pt 9):2271-9.

Kroll J. S., Wilks K. E., Farrant J. L., Langford P. R. 1998. Natural genetic exchange between *Haemophilus* and *Neisseria*: intergeneric transfer of chromosomal genes between major human pathogens.Proc Natl Acad Sci U S A.13;95(21):12381-5.

Kryazhimskiy S., Plotkin J. B. 2008. The population genetics of *dN/dS*.PLoS Genet. 4(12):e1000304.

Kwara A., Adegbola R.A., Corrah P.T., Weber M., Achtman M., Morelli G., Caugant D.A., Greenwood B.M. 1998. Meningitis caused by a serogroup W135 clone of the ET-37 complex of *Neisseria meningitidis* in West Africa. Trop Med Int Health. 3(9):742-6.

Lacks S. 1979. Uptake of circular deoxyribonucleic acid and mechanism of deoxyribonucleic acid transport in genetic transformation of *Streptococcus pneumoniae*. J. Bacteriol. 138:404–9.

Langford P. R., Loynds B. M., Kroll J. S. 1992. Copper-zinc superoxide dismutase in *Haemophilus* species. J Gen Microbiol. 138(3): 517-22.

Lemey, 2009. The Phylogenetic Handbook.4th Ed. Cambridge University Press.

Levine J.S, Strauss N. 1965. Lag periodcharacterizing the entry of transforming deoxyribonucleic acid into *Bacillus subtilis*. J. Bacteriol. 89:281–87

Lo H., Tang C. M., Exley R. M. 2009. Mechanisms of avoidance of host immunity by *Neisseria meningitidis* and its effect on vaccine development.Lancet Infect Dis.9(7):418-27.

Lorenz M.G, Wackernagel W. 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. 58: 563–602

Love P.E, Lyle M.J, Yasbin R.E. 1985. DNAdamage-inducible (din) loci are transcriptionally activated in competent *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 82: 6201–5.

MacFadyen, L. P. 2001. Competence development by Haemophilus influenzae is regulated by the availability of nucleic acid precursors. Mol. Microbiol. 40, 700–707.

Marrs C.F., Krasan G.P., McCrea K.W.2001. *Haemophilus influenza*- human specific bacteria. Front Biosci 6:41-60.

McCord J. M. and Fridovich I. 1969. Superoxide dismutase. An enzymatic function for erythrocuprein (hemocuprein). J Biol Chem. 224, 6049-6055.

McCrea K. W., Xie J., LaCross N., Patel M., Mukundan D., Murphy T. F., Marrs C. F., Gilsdorf J. R. 2008. Relationships of nontypeable *Haemophilus influenzae* strains to hemolytic and nonhemolytic *Haemophilus haemolyticus* strains.J Clin Microbiol. 46(2):406-16.

McCrea K. W, Wang M. L., Xie J., Sandstedt S. A., Davis G. S., Lee J. H. , Marrs C. F., Gilsdorf J. R. 2010. Prevalence of the sodC gene in nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus* by microarray-based hybridization.J Clin Microbiol. 48(3):714-9.

Mes T. H., Van Putten J. P. 2007. Positively selected codons in immune-exposed loops of the vaccine candidate OMP-P1 of *Haemophilus influenzae*.J Mol Evol. 64 (4): 411-22.

Medina, DB. 2010. Atividade Peroxidásica da Enzima Superóxido Dismutase 1 Humana: Produção do radical carbonato, Dimerização covalente da enzima e Implicações para a Esclerose Lateral Amiotrófica. Tese de Doutorado, Instituto de Química, Universidade de São Paulo, USP.

Morello JA.Smaron MF, Saxon E, Wood L, McCarthy C. 1991. Diagnosis of measles by fluorescent antibody and culture of nasopharyngeal secretions.J Virol Methods. Jun;33(1-2):223-9.

Morrison DA, Guild WR. 1973. Structure of deoxyribonucleic acid on the cell surface during uptake by *Pneumococcus*. J. Bacteriol. 115:1055–62.

Moxon E.R.1985 The molecular basis of *Haemophilus influenzae* virulence. J R Coll Physicians Lond 19:174-8.

Moxon E.R., Kroll J.S. 1988. Type b capsular polysaccharide as a virulence factor of *Haemophilus influenzae*. Vaccine 6:113-5.

Musser JM, Kroll JS, Moxon ER, Selander RK. 1988. Evolutionary genetics of the encapsulated strains of Haemophilus influenzae. Proc Natl Acad Sci U S A. Oct;85(20):7758-62.

Nassif X., Bourdoulous S., Eugène E., Couraud P.O. 2002. How do extracellular pathogens cross the blood-brain barrier. Trends Microbiol. 10(5):227-32. Review.

Nielsen, R.; Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 148: 929-936

Norrod P. and Morse S. A. 1979. Absence of superoxide dismutase in soe strains of *Neisseria Gonorrhoeae.* Biochem Biophys Res Commun. 90, 1287-1294.

Palmen R, Vosman B, Buijsman P, Breek CK, Hellingwerf KJ. 1993. Physiological characterization of natural transformation in *Acinetobacter calcoaceticus*. J. Gen. Microbiol. 139:295–305

Parker M. W. and Blake C. C. 1988. Iron and manganese-containing superoxide dismutases can be distinguished by analysis of their primary structures. FEBS Lett. 229: 377-382.

Piechowska M, Fox MS. 1971. Fate of transforming deoxyribonuclease in *Bacillus subtilis*. J. Bacteriol. 108:680–89

Pollard SJ, Farmer JG, Knight DM, Young PJ. 2002 Matrix effects in applying mono- and polyclonal ELISA systems to the analysis of weathered oils in contaminated soil. Environ Pollut.;117(1):5-8.

Provvedi R, Dubnau D. 1998. ComEA is a DNA receptor for transformation of competent *Bacillus subtilis*. Mol. Microbiol. In press

Puget K. and Michelson A. M. 1974. Isolation of a new copper-containing superoxide dismutase bacteriocuprein. Biochem Biophys Res Commun. 58, 830-838.

Redfield RJ. 1993. Evolution of natural transformation: testing the DNA repair hypothesis in *Bacillus subtilis* and *Haemophilus influenzae*. Genetics 133: 755–61

Redfield, RJ; Schrag MR; Dean, AM. 1997. The evolution of bacterial transformation: sex with poor relations. Genetics 146: 27-38.

Rosenstein N.E., Perkins B.A., Stephens D.S., Popovic T., Hughes J.M. 2001. Meningococcal disease. N Engl J Med. 3;344(18):1378-88. Review.

Sambrook J., Fritsch E. F., Maniatis T. 1989. Molecular Cloning – A Laboratory Manual. 2nd edition. Cold Spring Harbor Laboratory Press, USA.

Smith HO, Tomb JF, Dougherty BA, Fleischmann RD, Venter JC. 1995. Frequency and distribution of DNA uptake signal sequences in the *Haemophilus influenzae* Rd genome. Science 269: 538–40.

Spruyt, A. 1994 Standardized protocols for condom breakage and slippage trials: a proposal. Am J Public Health Dec;84(12):1897-900.

Stabel T. J., Sha Z. and Mayfield J. E. 1994. Periplasmic location of *Brucella abortus* Cu/Zn superoxide dismutase. Vet. Microbiol. 38, 305-314.

Steinman H. M. 1982. Copper-zinc superoxide dismutase from *Caulobacter crescentus* CB15. J Biol Chem. 257, 10283-10293.

Steinman H. M. 1985. Bacteriocuprein superoxide dismutases in *Pseudomonads.* J Bacteriol. 162, 1255-1260

Steinmoen, H.; Knutsen, E; Havarstein, L. 2002 S. Induction of natural competence in Streptococcus pneumoniae triggers lysis and DNA release from a subfraction of the cell population. Proc. Natl Acad. Sci. USA 99, 7681–7686.

Steinmoen, H.; Teigen, A; Havarstein, L. S. 2003 Competenceinduced cells of Streptococcus pneumoniae lyse competence-deficient cells of the same strain during cocultivation. J. Bacteriol. 185, 7176–7183 .

Stephens D., Greenwood B., Brandtzaeg P. 2007. Epidemic meningitis, meningococcaemia, and *Neisseia meningitis*. Lancet 369: 2196-2206.

Swartley J.S., Marfin A.A., Edupuganti S. 1997. Capsule switching of *Neisseria meningitidis*. Proc. Natl. Acad. Sci. USA 94: 271-276.

Taha MK. 2000. Simultaneous approach for nonculture PCR-based identification and serogroup prediction of Neisseria meningitidis.J Clin Microbiol. Feb;38(2):855-7.

Taha MK, Achtman M, Alonso JM, Greenwood B, Ramsay M, Fox A, Gray S, Kaczmarski E. 2000 Serogroup W135 meningococcal disease in Hajj pilgrims.
Lancet. Dec 23-30;356(9248):2159.

Taha M.K., Deghmane A.E., Antignac A. 2002. The duality of virulence and transmissibility in *Neisseria meningitidis*. TRENDS in Microbiol. 1: 376-382.

Tettelin H., Saunders N.J., Heidelberg J.F., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R.T., Peterson J.D., Hickey E.K., Haft D.H.,Salzberg S.L., White O. **Venter J.C. 2000** Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58**.** Science 287:1809-1815(2000)

Tortora G. J.; Funke, B. R. e Case C. L. 2000. Microbiologia**.** 6 ed, Artmed, Porto Alegre.

Tzeng Y.L. and Stephens D.S.2000. Epidemiology and pathogenesis of *Neisseria meningitidis*.Microbes Infect. 2(6):687-700. Review.

Urwin R., Holmes E. C., Fox A. J., Derrick J. P., Maiden M. C. 2002. Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB.Mol Biol Evol. 19(10):1686-94.

Van Camp W.,Bowler C., Villarroel R., Tsang E. W., Van Montagu M., Inzé D. 1990. Characterization of iron superoxide dismutase cDNAs from plants obtained by genetic complementation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 87: 9903-9907.

Van Deuren M., Brandtzaeq P., Van der Meer J. M. 2000. Update on meningococcal disease with emphasis on pathogenesis and clinical management. Clin Microbiol Rev 13 (1): 144-166.

Wilks K.E., Dunn K.L., Farrant J.L., Reddin K.M., Gorringe A.R., Langford P.R., Kroll J.S. 1998 Periplasmic superoxide dismutase in meningococcal pathogenicity.Infect. Immun. 66:213-217.

Yang, Z. 1997. PAML: a program for package for phylogenetic analysis by maximum likelihood. CABIOS 15: 555-556

Yang, Z.; R. Nielsen. 1998. Synonymous and nonsynonymous rate variation in nuclear genes of mammals. Journal of Molecular Evolution, 46: 409-418

Yang, Z.; R. Nielsen; N. Goldman. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155: 431-449

Yang, Z. 2002. Accuracy and power of Bayes prediction of amino acid sites under positive selection. Molecular Biology and Evolution 19: 950-958.

Yang, Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. Journal of Molecular Evolution 59:121-13.

Yasdankhah. S. P.; Caugant, D. A. 2004. *Neisseria meningitidis* : an overview of the carriage state.J Med Microbiol. 53(Pt 9): 821-32. Review.