

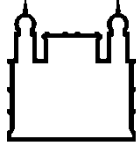
MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ

Doutorado pelo Programa de Pós-Graduação  
em Biologia Computacional e Sistemas

MINERAÇÃO DE DADOS METAGENÔMICOS VISANDO A  
ANÁLISE DA DIVERSIDADE DE SERINA B-LACTAMASES EM  
DIFERENTES AMBIENTES

ADRIANA MACHADO FRÓES

Rio de Janeiro  
Junho de 2013



Ministério da Saúde

FIOCRUZ  
Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

*ADRIANA MACHADO FRÓES*

Mineração de Dados Metagenômicos Visando a Análise da Diversidade de Serina  $\beta$ -Lactamases em Diferentes Ambientes

Tese apresentada ao Instituto Oswaldo Cruz  
como parte dos requisitos para obtenção do título  
de Doutor em Biologia Computacional e Sistemas

**Orientador:** Prof. Dr Alberto Martin Rivera Dávila

RIO DE JANEIRO

Junho de 2013

Ficha catalográfica elaborada pela  
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

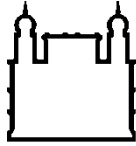
F926 Fróes, Adriana Machado

Mineração de dados metagenômicos visando a análise da diversidade de serina b-lactamases em diferentes ambientes / Adriana Machado Fróes. – Rio de Janeiro, 2013.  
xviii, 184 f. : il. ; 30 cm.

Tese (Doutorado) – Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2013.  
Bibliografia: f. 150-164

1. Beta-lactamases. 2. Metagenômica. 3. Esgotos hospitalares. 4. Comunidade microbiana. I. Título.

CDD 615.329



Ministério da Saúde

FIOCRUZ  
Fundação Oswaldo Cruz

**INSTITUTO OSWALDO CRUZ**  
**Pós-Graduação em Biologia Computacional e Sistemas**

***ADRIANA MACHADO FRÓES***

**MINERAÇÃO DE DADOS METAGENÔMICOS VISANDO A ANÁLISE DA  
DIVERSIDADE DE DE SERINA  $\beta$ -LACTAMASES EM DIFERENTES AMBIENTES**

**ORIENTADOR: Dr. Alberto Martin Rivera Dávila**

**Aprovada em: 27/ 06 / 2013**

**EXAMINADORES:**

**Prof. Dr. Oswaldo Gonçalves Cruz (FIOCRUZ/PROCC) - Presidente**

**Prof. Dra. Ana Paula D'Alincant Carvalho Assef (FIOCRUZ/IOC)**

**Prof. Dr. Fabio Faria da Mota (FIOCRUZ/IOC)**

**Prof. Dr. Juliano de Carvalho Cury (UFSJ)**

**Prof. Dra. Maria Claudia Calvacanti (IME)**

Rio de Janeiro, 27 de junho de 2013

## AGRADECIMENTOS

Aos meus queridos e amados pais Berenice e Guilherme por todo amor, carinho, apoio psicológico e financeiro, incentivo e pela paciência nos últimos e grandes momentos de estresse.

À Luiza por ser uma irmã maravilhosa e ter me apoiado muito nos momentos mais difíceis durante essa jornada.

Ao meu orientador Dr. Alberto Dávila pela oportunidade, aprendizado, pelo auxílio nesse período de desenvolvimento da tese e por ajudar a encaminhar minha conduta na área científica.

À Dra Yara Traub-Cseko por ter disponibilizado o espaço de seu laboratório de Parasitos e Vetores (IOC/FIOCRUZ), os materiais e equipamentos necessários para as etapas de biologia molecular da tese.

Ao Dr. Jaime de Oliveira e sua equipe (Marcelo, Marcos e Jane), responsáveis pelas análises físico-químicas.

Aos dois hospitais públicos que possibilitaram o desenvolvimento deste trabalho, em especial aos contatos que permitiram as coletas.

Aos colegas de laboratório Rodrigo Jardim, Fábio Mota, Diogo Tschoeke, Rafael Cuadrat e Fábio Bernardo pelo constante apoio e ajuda nas dúvidas e discussões que sempre me ajudaram muito, além de terem contribuído para o avanço de etapas importantes da parte “in silico”, contribuindo com ideias e discussões que ajudaram a engrandecer este trabalho, incluindo os momentos de descontração e amizade. À Gisele Lopes, Joana Lima e Aline Dumaresq por terem me ajudado na parte laboratorial, além das intermináveis discussões de trabalho e conversas que sempre foram muito enriquecedoras, incluindo todo apoio, carinho e amizade.

Aos amigos Tiana, Luciana, Daniela, Amanda, Milene, Daniel, Ana Carolina, Kele, Ana Bahia, Fernando e aos amigos do laboratório 304 e do laboratório 201, incluindo também todo o pessoal do laboratório 301 pela amizade e apoio.

Ao Felipe Figueiredo pelo incentivo constante, pelo apoio, pela amizade na maior parte desta etapa, além da grande ajuda nas discussões sobre a tese e ao me questionar sobre várias etapas da metodologia que me fizeram refletir bastante e contribuíram para a escrita da tese.

À amiga Nicole Scherer, pelo apoio, amizade e por ter me ajudado muito na etapa final, incluindo as discussões e sugestões que contribuíram para a conclusão da dissertação do trabalho.

Por todos os outros amigos que não mencionei, mas também ajudaram de forma direta ou indiretamente nesse período, como muito carinho, apoio e paciência.

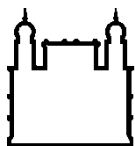
À amiga Cintia Palu por ter contribuído bastante com a formatação da tese.

À Plataforma Genômica - Sequenciamento de DNA - PDTIS/FIOCRUZ (RPT01A) pela imensa ajuda na preparação das amostras e posterior sequenciamento.

À FIOCRUZ (incluindo o IOC) pelo auxílio financeiro durante o período do doutorado e conclusão da tese.

À Pós-Graduação da Biologia Computacional e Sistemas pelo auxílio financeiro aos congressos que sempre ajudaram muito na troca de conhecimentos e aprendizagem, contribuindo para o meu amadurecimento científico.

Aos doutores membros da banca Doutores Oswaldo Cruz, Ana Paula Assef, Fabio Mota, Maria Cláudia Cavalcanti e Juliano de Carvalho Cury pela participação e sugestões realizadas para a melhoria da tese.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

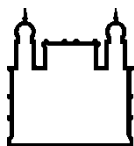
### MINERAÇÃO DE DADOS METAGENÔMICOS VISANDO A ANÁLISE DA DIVERSIDADE DE SERINA $\beta$ -LACTAMASES EM DIFERENTES AMBIENTES

#### RESUMO

#### TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

**Adriana Machado Fróes**

O uso inadequado de antibióticos tem causado um grande impacto na saúde pública devido à seleção e aumento de bactérias patogênicas multirresistentes e de difícil tratamento. Dentre os antibióticos administrados na área clínica, os  $\beta$ -lactâmicos são os mais usados e atuam no bloqueio da síntese da parede celular bacteriana, levando à sua morte. Bactérias desenvolveram resistência a esses antibióticos, devido, inclusive, à produção de  $\beta$ -lactamases que são capazes de clivar o anel  $\beta$ -lactâmico, permitindo que a síntese da parede celular bacteriana ocorra. Os hospitais são uma grande fonte de seleção de bactérias multirresistentes, passando para os esgotos municipais e ambientes, mas pouco se sabe a respeito da prevalência e diversidade desses genes em esgotos hospitalares. Sendo assim, o presente trabalho teve como objetivo a exploração taxonômica e funcional do metagenoma dos esgotos de dois hospitais do Rio (EHRJ), assim como a análise da diversidade de  $\beta$ -lactamases presentes nesses hospitais e em bases de dados metagenômicos públicas (CAMERA e IMG/M). A análise taxonômica das sequências dos esgotos hospitalares do Rio revelou maior abundância de Firmicutes (~55%) e Bacteroidetes (~30%) em ambas as amostras hospitalares. A mineração de  $\beta$ -lactamases foi realizada através de buscas com perfis HMM para cada classe (A, C e D) em bases de dados metagenômicos, incluindo a de EHRJ. Os resultados indicaram a presença de prováveis homólogos de  $\beta$ -lactamases nos diferentes ambientes, tendo sua anotação funcional confirmada por similaridade com bases de domínios conservados e filogenia, mostrando ser um método sensível e eficaz. Em relação às amostras de esgotos hospitalares, a filogenia mostrou que essas sequências foram agrupadas próximas às  $\beta$ -lactamases representativas dos grupos Firmicutes e Bacteroidetes, corroborando com os resultados de análise taxonômica. As sequências das amostras EHRJ formaram um clado entre si, na maioria das vezes, distantes das sequências dos diferentes projetos metagenômicos previamente estudados de outros ambientes que agruparam próximas de sequências representativas do grupo Proteobacteria, em sua maioria. O hospital da Zona Sul do Rio apresentou uma maior abundância de  $\beta$ -lactamases da classe A (50%), enquanto no hospital da Zona Norte a classe D (55%). Em relação às subclasses de  $\beta$ -lactamases, ACC (53%), CfxA (65%), OXA-10 (30%) e LCR-1 (27%) foram as subclasses de  $\beta$ -lactamases mais abundantes detectadas no efluente da Zona Sul, enquanto nas amostras da Zona Norte as mais abundantes foram CfxA (28,4%) e CEPA (20%), ACC e FOX (~20%) e OXA-10 (32%). Essa diferença na distribuição dessas subclasses pode estar relacionada com a quantidade e tipos de antibióticos  $\beta$ -lactâmicos administrados em cada hospital. A análise taxonômica revelou que Firmicutes (~55%) e Bacteroidetes (~30%) foram os grupos mais abundantes presentes nas amostras de esgotos hospitalares do Rio.



Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

## **INSTITUTO OSWALDO CRUZ**

### **MINING METAGENOMIC DATA AIMING THE DIVERSITY ANALYSIS OF SERINE $\beta$ -LACTAMASES IN DIFFERENT ENVIRONMENTS**

#### **ABSTRACT**

#### **PHD THESIS IN BIOLOGIA COMPUTACIONAL E SISTEMAS**

**Adriana Machado Fróes**

The misuse of antibiotics has caused a huge impact on public health because of the selection and increase of multiresistant pathogenic bacteria, making its treatment extremely hard. Among the antibiotics taught in the clinical area,  $\beta$ -lactamics are the most used and acts on blocking the bacteria cell wall synthesis, causing its death. However, bacteria have evolved resistance to these antibiotics due mainly to the production of  $\beta$ -lactamases which are capable of cleaving the  $\beta$ -lactam ring, allowing the cell wall synthesis to occur. Hospitals are a great source of multiresistant bacteria selection, and it passes to municipal and environments effluents. However, little is known about the real prevalence of these genes in hospitals sewage. This work applied high-throughput sequencing and metagenomic analyses to explore the diversity and functional analysis in bacterial community from two sewage hospitals in Rio de Janeiro-RJ and also to study the diversity of  $\beta$ -lactamases in those two hospitals environments, including those available in public databases such as CAMERA and IMG/M. Searches with pHMMs were performed against metagenomes datasets using appropriate profiles for studies concerning  $\beta$ -lactamases diversity. Besides, functional and taxonomic characterization of  $\beta$ -lactamases was performed using samples from the two hospitals sewage, including phylogenetic analysis of the possible  $\beta$ -lactamases. Results from the searches with pHMMs showed that these profiles were capable of mining  $\beta$ -lactamases from different sources, checked by similarity analysis with conserved patterns databases and phylogeny showing the efficacy and sensitivity of our pHMM-based approach in detecting putative  $\beta$ -lactamases sequences. Phylogeny assisted to classify these enzymes showing the sewage hospitals probable  $\beta$ -lactamases grouped along the same clade and close to Firmicutes and Bacteroidetes sequences, and distant from probable  $\beta$ -lactamases retrieved from metagenomic projects from environmental (especially aquatic and terrestrial), engineered and host-associated ecosystems, which grouped closer to  $\beta$ -lactamases from Proteobacteria, mainly. Hospital from South Zone (SZ) of Rio presented a greater abundance of class A  $\beta$ -lactamases (50%), while in the hospital from North Zone (NZ) were class D (55%). Concerning the subclasses of  $\beta$ -lactamases, ACC (53%), CfxA (65%), OXA-10 (30%) and LCR-1 (27%) were the most abundant detected on sewage samples from the SZ hospital of Rio as on samples from NZ hospital the most abundant were CfxA (28,4%) and CEPA (20%), ACC and FOX (~20%) and OXA-10 (32%). Such difference in this subclasses distributions could be related to the quantities and types of  $\beta$ -lactams antibiotics administered in the two hospitals. Taxonomic analysis revealed that Firmicutes (~55%) and Bacteroidetes (~30%) were the most abundant groups composing the hospital sewage samples from Rio.



# SUMÁRIO

Agradecimentos .....	v
Índice de Figuras.....	xiii
Lista de Tabelas.....	xviii
1 INTRODUÇÃO .....	20
1.1 Água, poluição e esgoto hospitalar.....	20
1.2 Resistência bacteriana a antibióticos.....	22
1.2.1 Resistoma .....	22
1.2.2 $\beta$ -lactâmicos .....	23
1.3 Alvo dos $\beta$ -lactâmicos.....	25
1.3.1 Parede celular bacteriana: composição geral .....	26
1.3.2 Parede celular bacteriana: Gram-positivas e Gram-negativas .....	27
1.4 $\beta$ -lactamases .....	28
1.4.1 Classificação das $\beta$ -lactamases .....	29
1.4.2 Mecanismo de atividade.....	31
1.4.3 Filogenia e características moleculares .....	31
1.4.4 Especificações sobre cada classe de serina- $\beta$ -lactamase .....	33
1.4.5 Sítio ativo das PBPs e serina- $\beta$ -lactamases.....	35
1.5 Reconhecimento da função enzimática por similaridade de sequências .....	38
1.6 Metagenômica .....	39
1.7 Bases de dados metagenômicos e de resistência a antibióticos .....	43
2 JUSTIFICATIVA .....	45
3 OBJETIVOS .....	47
3.1 Geral.....	47
3.2 Específicos .....	47
4 METODOLOGIA.....	48
4.1 Escolha dos locais de coleta .....	48
4.2 Coleta da amostra (efluente hospitalar).....	48
4.3 Análise físico-química e colimetria.....	50
4.4 Extração de DNA metagenômico.....	50
4.5 Caracterização do DNA metagenômico.....	51
4.5.1 Amplificação de uma região do gene ribossomal 16S.....	52

4.5.2	Clonagem das bibliotecas de 16S ribossomal e sequenciamento pelo método de Sanger .....	52
4.5.3	Pré-processamento das sequências de 16S ribossomal geradas pelo método de Sanger .....	53
4.5.4	Estimativa de riqueza e diversidade da comunidade bacteriana utilizando parte do gene ribossomal 16S .....	53
4.5.5	Classificação taxonômica .....	54
4.5.6	Pirosequenciamento das amostras .....	54
4.5.7	Pré-processamento das sequências geradas pelo aparelho 454/Life Science .....	55
4.6	Construção de pHMMs para a mineração de homólogos de $\beta$ -lactamases..	55
4.6.1	Seleção de sequências “sementes” para a construção de alinhamentos.....	55
4.6.2	Alinhamento múltiplo de sequências .....	55
4.6.3	Perfis de Cadeia Oculta de Markov (pHMM) .....	56
4.6.4	Construção dos perfis de Cadeia Oculta de Markov (pHMM) .....	57
4.7	Busca por homólogos de serina $\beta$ -lactamase em bases de dados metagenômicos.....	58
4.7.1	Bases de estudo.....	58
4.7.2	Busca por homólogos de $\beta$ -lactamases .....	59
4.8	Anotação funcional através da presença de domínios conservados por similaridade.....	59
4.9	Construção de pHMM de regiões específicas de $\beta$ -lactamases .....	60
4.10	Busca por homólogos de serina $\beta$ -lactamases em bases de esgotos hospitalares.....	61
4.11	Cálculo da abundância relativa dos homólogos de serina $\beta$ -lactamases ..	61
4.12	Análise da diversidade filogenética dos homólogos de serina $\beta$ -lactamases em bases ambientais e obtidas por pirosequenciamento .....	62
4.13	Caracterização dos metagenomas hospitalares do Rio de Janeiro.....	64
4.14	. Curva de rarefação.....	65
5	RESULTADOS .....	67
5.1	Análise físico-química e contagem de coliformes totais e fecais das amostras de EHRJ.....	67
5.2	Extração e quantificação do DNA metagenômico dos EHRJ .....	67
5.3	Caracterização do DNA metagenômico dos EHRJ.....	68

5.3.1	Amplificação por PCR da região do 16S ribossomal, clonagem da biblioteca e sequenciamento pelo método de Sanger .....	68
5.3.2	Estimativa da diversidade e riqueza de UTOs .....	68
5.3.3	Informações sobre os dados gerados pelo pirosequenciamento das amostras de EHRJ.....	69
5.4	Pré-processamento das sequências geradas pela plataforma 454/Life Science .....	73
5.5	Construção de pHMMs de serina $\beta$ -lactamases .....	74
5.6	Busca por homólogos de serina $\beta$ -lactamases .....	77
5.7	Anotação funcional das sequências homólogas às serina $\beta$ -lactamases .....	78
5.8	Abundância relativa de $\beta$ -lactamases .....	81
5.8.1	Em amostras de esgotos hospitalares .....	81
5.8.2	Em projetos metagenômicos da base CAMERA.....	87
5.9	Análise filogenética das sequências ambientais homólogas às serina $\beta$ -lactamases.....	89
5.9.1	Classe A.....	89
5.9.2	Classe C .....	94
5.9.2	Classe D.....	97
5.10	Caracterização do metagenoma de esgotos hospitalares do Rio de Janeiro.....	100
5.10.1	Análise dos dados gerados a partir da amplificação do 16S ribossomal bacteriano.....	100
5.10.2	Dados metagenômicos dos EHRJ obtidos a partir do pirosequenciamento.....	102
5.11	Perfil do potencial metabólico do metagenoma .....	122
6	DISCUSSÃO .....	130
6.1	Parâmetros físico-químicos e colimetria das amostras de esgotos hospitalares do Rio de Janeiro.....	130
6.2	Informações sobre os dados gerados pelo pirosequenciamento das amostras de EHRJ.....	132
6.3	Caracterização do metagenoma dos esgotos hospitalares do Rio de Janeiro.....	132
6.3.1	Perfil taxonômico da comunidade bacteriana presente nas amostras de esgotos dos hospitais da Zona Norte e Zona Sul .....	132

6.4	Diversidade de $\beta$ -lactamases em esgotos hospitalares e em projetos de metagenomas públicos .....	136
6.5	Potencial metabólico da comunidade bacteriana presente nas amostras de esgotos dos hospitais da Zona Norte e Zona Sul .....	147
6.6	Considerações finais .....	149
6.7	Perspectivas futuras .....	150
7	CONCLUSÕES .....	152
8	REFERÊNCIA BIBLIOGRÁFICA.....	154
	APÊNDICE .....	169
	ANEXOS.....	183

## ÍNDICE DE FIGURAS

<b>Figura 1.1.</b> Estrutura química de antibióticos $\beta$ -lactâmicos: (1) penicilina, (2) cefalosporina, (3) monobactâmico e (4) carbapenem.....	24
<b>Figura 1.2.</b> Esquema da composição da molécula de peptidoglicano de uma bactéria Gram-negativa, a <i>E. coli</i> . M: N-acetil murâmico, G: N-acetil glicosamina....	27
<b>Figura 1.3.</b> Esquema da composição da parede celular de bactérias Gram-positivas e Gram-negativas e a disposição de seus componentes.....	28
<b>Figura 1.4.</b> Reação enzimática das $\beta$ -lactamases (3.2.5.8).....	31
<b>Figura 1.5.</b> Árvore filogenética das três classes de serina $\beta$ -lactamases, a partir do alinhamento estrutural, junto com seu homólogo, as DD-peptidases.....	32
<b>Figura 1.6.</b> Estrutura molecular do ancestral das serina $\beta$ -lactamases, da enzima homóloga DD-peptidase R61 e das $\beta$ -lactamases classe A (TEM), classe C (AmpC) e classe D (OXA).....	36
<b>Figura 4.1</b> Box 1: Base de dados de proteínas de onde foram retiradas as sequências de aminoácidos das $\beta$ -lactamases das classes A, C e D para a construção dos pHMMs de cada classe. Box 2: Alinhamento múltiplo das sequências de proteína das classes A, C e D de $\beta$ -lactamases com o programa MAFFT para posterior construção dos pHMMs referentes a cada classe.....	56
<b>Figura 4.2.</b> Esquema da metodologia relacionada aos itens 4.6.2, 4.6.3 e 4.7.2.....	59
<b>Figura 4.3.</b> Fluxograma resumido da metodologia referente à mineração e análise da diversidade de serina $\beta$ -lactamases dispostas em projetos metagenômicos e esgotos hospitalares do Rio de Janeiro.....	64
<b>Figura 4.4.</b> Fluxograma resumido da metodologia descrita neste trabalho, incluindo a parte experimental e a parte “In silico”, que envolve também as análises da parte experimental.....	66
<b>Figura 5.1.</b> Distribuição do comprimento das sequências do pirosequenciamento das réplicas 01 (A) e 02 (B) coletadas em um hospital público da Zona Sul do Rio de Janeiro (ZS), após serem filtradas por qualidade, redundância e tamanho.....	70
<b>Figura 5.2.</b> Distribuição do comprimento das sequências do pirosequenciamento das réplicas 01 (A) e 02 (B) coletadas em um hospital público da Zona Norte do Rio de Janeiro (ZN), após serem filtradas por qualidade, redundância e tamanho.....	72

<b>Figura 5.3.</b> Imagem ilustrativa do logo parcial do pHMM dos principais motivos da $\beta$ -lactamase classe A, gerado pelo programa LogoMat-M que se encontram nas caixas 1,2 e 3.....	75
<b>Figura 5.4.</b> Imagem ilustrativa do logo parcial do pHMM dos principais motivos da $\beta$ -lactamase classe C, gerado pelo programa LogoMat-M. que se encontram nas caixas 1,2 e 3.....	76
<b>Figura 5.5.</b> Imagem ilustrativa do logo parcial do pHMM dos principais motivos da $\beta$ -lactamase classe D, gerado pelo programa LogoMat-M. que se encontram nas caixas 1,2 e 3.....	77
<b>Figura 5.6.</b> Abundância relativa (%) dos prováveis homólogos das três classes de serina $\beta$ -lactamases relacionadas às amostras de esgotos hospitalares ZS e ZN.....	82
<b>Figura 5.7.</b> Frequência relativa das subclasses das $\beta$ -lactamases classe A presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.....	83
<b>Figura 5.8.</b> Frequência relativa (%) das subclasses das $\beta$ -lactamases classe C presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.....	85
<b>Figura 5.9.</b> Frequência relativa (%) das subclasses da $\beta$ -lactamase classe D presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.....	86
<b>Figura 5.10.</b> Frequência relativa (%) dos prováveis homólogos de serina $\beta$ -lactamases classe A, C e D, minerados a partir dos pHMM, presentes nos diferentes projetos presentes na base CAMERA.....	88
<b>Figura 5.11.</b> Filogenia das prováveis $\beta$ -lactamases da classe A, recuperadas da base de dados de esgotos hospitalares ZS e ZN.....	90
<b>Figura 5.12.</b> Filogenia das prováveis $\beta$ -lactamases da classe A, recuperadas das bases de dados de projetos de metagenoma (CAMERA e IMG/M) que estão representadas com um triângulo fechado verde.....	93
<b>Figura 5.13.</b> Filogenia das prováveis $\beta$ -lactamases da classe C, recuperadas das bases de dados metagenômicos CAMERA e IMG/M, e de esgotos hospitalares de amostras coletadas nos hospitais públicos da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro.....	96
<b>Figura 5.14.</b> Filogenia das prováveis sequências de $\beta$ -lactamases da classe D, recuperadas das bases de dados metagenômicos CAMERA, IMG/M e dos	

esgotos hospitalares de amostras coletadas nos hospitais públicos da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro.....	99
<b>Figura 5.15.</b> Curva de rarefação. Sequências parciais do gene 16S ribossomal (rrs) de Bacteria foram utilizadas para estimar a diversidade bacteriana nas amostras de esgotos hospitalares do Rio, ZS e ZN, baseado na identidade de sequências de 97% como ponto de corte para o delineamento de unidades taxonômicas operacionais (UTOs) pelo MOTHUR.....	100
<b>Figura 5.16.</b> Abundância relativa (%) dos filós mais predominantes nas amostras de esgotos hospitalares da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro a partir da amplificação do gene ribossomal 16S.....	101
<b>Figura 5.17.</b> Curva de rarefação das amostras ZN e ZS, localizados no Rio de Janeiro-RJ, gerada a partir do MEGAN, de acordo com as sequências pertencentes ao grupo Bacteria que foram classificadas ao nível de gênero, a partir dos resultados do BLASTN.....	102
<b>Figura 5.18.</b> Diversidade taxonômica das amostras de esgotos hospitalares ZS e ZN localizados no Rio de Janeiro-RJ.....	103
<b>Figura 5.19.</b> Gráfico da distribuição taxonômica de Bacteria presente nas amostras dos esgotos hospitalares da Zona Sul (ZS01, ZS02) e da Zona Norte (ZN01 e ZN02).....	104
<b>Figura 5.20.</b> Gráfico em barras de cada réplica das amostras hospitalares gerado pelo programa MEGAN 4. ....	105
<b>Figura 5.21.</b> Distribuição taxonômica (gráfico) e frequência relativa (tabela) do filo Firmicutes presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro, RJ.....	106
<b>Figura 5.22.</b> Distribuição taxonômica do grupo Clostridiales presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro, RJ.....	107
<b>Figura 5.23.</b> Distribuição taxonômica das sequências pertencentes à família Ruminococcaceae, presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ.....	109
<b>Figura 5.24.</b> Distribuição taxonômica das sequências pertencentes à família Lachnospiraceae, presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	110
<b>Figura 5.25.</b> Distribuição taxonômica (gráfico) e valores da frequência relativa (tabela) do filo Bacteroidetes presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	112

<b>Figura 5.26.</b> Distribuição taxonômica da ordem Bacteroidales presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	113
<b>Figura 5.27.</b> Abundância relativa dos gêneros do filo Bacteroides presentes nas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	114
<b>Figura 5.28</b> Distribuição taxonômica do filo Proteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	115
<b>Figura 5.29</b> Distribuição taxonômica da classe Alphaproteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ...	116
<b>Figura 5.30.</b> Distribuição taxonômica (gráfico) e valor da frequência relativa (tabela) da classe Betaproteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ.....	117
<b>Figura 5.31.</b> Distribuição taxonômica da classe Gammaproteobacteria presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ.....	118
<b>Figura 5.32.</b> Abundância relativa dos gêneros da família Enterobacteriaceae presentes nas amostras dos esgotos hospitalares localizados no Rio de Janeiro, RJ.....	119
<b>Figura 5.33.</b> Distribuição taxonômica da classe Delta-Proteobacteria presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ.....	120
<b>Figura 5.34.</b> Distribuição taxonômica dos gêneros pertencentes à Ordem Campylobacterales presentes nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ.....	122
<b>Figura 5.35.</b> Classificação funcional dos subsistemas presentes nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) utilizando a base SEED através do programa MEGAN 4.....	124
<b>Figura 5.36.</b> Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de carboidratos presente nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) pela base SEED presente no MEGAN4.....	125
<b>Figura 5.37.</b> Classificação funcional das proteínas relacionadas ao subsistema de virulência presente nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) pela base SEED presente no MEGAN4.....	126
<b>Figura 5.38</b> Classificação funcional das proteínas relacionadas ao subsistema de resistência a antibióticos e compostos tóxicos presentes nas réplicas 01 das	



amostras dos hospitais da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo) pela base SEED presente no MEGAN4.....127

**Figura 5.39** Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de proteínas presentes nas réplicas 01 das amostras hospitalares da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo), pela base SEED presente no MEGAN4.....128

**Figura 5.40** Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de DNA presente nas réplicas 01 das amostras hospitalares da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo), pela base SEED presente no MEGAN4.....129

## LISTA DE TABELAS

<b>Tabela 1.1.</b> Principais regiões conservadas (motivos) referentes às classes de serina $\beta$ -lactamases.....	39
<b>Tabela 5.1.</b> Análise físico-química das amostras de esgotos de dois hospitais situados na Zona Sul e Zona Norte da cidade do Rio de Janeiro.....	68
<b>Tabela 5.2.</b> Estimativa de riqueza de UTOs e índices de diversidade calculados a partir de bibliotecas de DNA 16S de Bactéria de dois hospitais públicos do Rio de Janeiro.....	70
<b>Tabela 5.3.</b> Dados numéricos das sequências das réplicas 01 (ZS01) e 02 (ZS02) do hospital da Zona Sul do Rio de Janeiro, geradas após o pirosequenciamento com a plataforma 454/Life Science e filtradas por qualidade, redundância e tamanho.....	71
<b>Tabela 5.4.</b> Dados numéricos das sequências das réplicas 01 (ZN01) e 02 (ZN02) do hospital da Zona Norte do Rio de Janeiro, geradas após o pirosequenciamento com a plataforma 454/Life Science e filtradas por qualidade, redundância e tamanho.....	72
<b>Tabela 5.5.</b> Resultado dos números de sequências de prováveis homólogos de $\beta$ -lactamases obtidas após a busca com o “hmmsearch” com os perfis de cada classe (pré-filtro), e após o “hmmsearch” utilizando os perfis de regiões específicas de $\beta$ -lactamases (pós-filtro) contra as bases CAMERA e IMG/M.....	77
<b>Tabela 5.6.</b> Resultado dos números de sequências obtidas após a busca com o “hmmsearch” (pré-filtro) a partir dos perfis de cada classe de $\beta$ -lactamases (A, C e D) contra as bases de esgotos ZS e ZN. ZS: Hospital da Zona Sul. ZN: Hospital da Zona Norte.....	79

## LISTA DE ABREVIACÕES

DNA – Ácido desoxirribonucléico

NCBI – “National Center for Biotechnology Information” (Centro Nacional de Informação Biotecnológica).

Refseq – “NCBI Reference Sequence Database” (Base de dados de referência do NCBI).

ARDB – “Antibiotic Resistance Database” (Base de dados de genes de resistência a antibióticos).

GRAs – Gene de resistência a antibióticos

EGMs – Elementos genéticos móveis

ESBLs –  $\beta$ -lactamases de espectro estendido

PBPs – “Penicillin binding proteins” (Proteínas que se ligam à penicilina)

pHMM – “Profile Hidden Markov Models” (Perfis de cadeia oculta de Markov)

MEGAN – “MEtaGenome ANalyzer”

LCA – “Lowest Common Ancestor” (menor ancestral comum ou ancestral comum mais próximo).

UTO – Unidade Taxonômica Operacional

DQO – Demanda Química de Oxigênio

DBO – Demanda Bioquímica de Oxigênio

pH – Potencial hidrogeniônico

ZN – Hospital da Zona Norte

ZS – Hospital da Zona Sul

RaxML – “Randomized Axelerated Maximum Likelihood”

TrimAL – “A tool for automated alignment trimming”

LBCS- Laboratório de Biologia Computacional e Sistemas

EHRJ- Esgotos hospitalares do Rio de Janeiro

Q – Índice de qualidade de base usado pelo programa PHRED, onde  $Q = -10 \log_{10}P$ , ou seja, se a qualidade do Phred for 30 para uma base, as chances de essa base ter sido convertida (ou “chamada”) de forma incorreta são 1 em 1000.

# 1 INTRODUÇÃO

## 1.1 Água, poluição e esgoto hospitalar

O propósito primário para a exigência de qualidade da água é a proteção à saúde pública. Os critérios clássicos para análise da qualidade de água são baseados principalmente na presença de coliformes totais e fecais (Resolução CONAMA Nº 274, de 29 de novembro de 2000). Contudo, a presença de coliformes na água é apenas um possível indicativo da ação antrópica no ambiente, já que o habitat natural e o principal nicho das enterobactérias é o trato intestinal do homem e de outros animais endotérmicos. A quantificação de coliformes, embora seja uma abordagem rápida e de baixo custo, não traz qualquer tipo de informação sobre as características e o verdadeiro potencial patogênico destes micro-organismos (Ciampi *et al.*, 1987). Além do mais, outros patógenos humanos, como vírus entéricos e cistos de protozoários, não fazem parte do grupo dos coliformes e conseqüentemente não são detectados nestes testes.

A água é um dos principais veículos de transmissão de diversos patógenos. Sua poluição tem como origem várias fontes, como os esgotos industriais, os resíduos da agricultura e os esgotos domésticos e hospitalares. Dentre as diversas fontes de poluentes da água, os dejetos oriundos dos esgotos domésticos e hospitalares são potencialmente mais nocivos no que diz respeito à contaminação por micro-organismos patogênicos. Este quadro se torna mais crítico quando se leva em consideração a condição precária do sistema de saneamento básico no país, contribuindo para os problemas de saúde associados à qualidade da água disponível à população. As águas residuais e dos esgotos carregam grandes quantidades de matéria orgânica e micro-organismos, sendo frequentemente lançada, sem tratamento prévio, nos rios, lagos e no mar, o que constitui uma grave ameaça para a saúde das populações (Prado, 2007).

O uso indiscriminado e incorreto na administração de antibióticos vem contribuindo para o surgimento de bactérias patogênicas multiresistentes, tornando o tratamento de doenças, causadas por esses patógenos, extremamente difícil, custoso ou até impossível, levando a ocorrência de casos

fatais de infecção hospitalar (Fajardo *et al.*, 2008). Desta forma, a presença de fármacos em águas residuais, entre eles os antibióticos, tem sido alvo de questionamentos e de pesquisa nas áreas de saúde pública e ambiental devido, principalmente, ao aumento da persistência de bactérias multiresistentes no meio ambiente.

Os esgotos hospitalares geralmente contêm altas concentrações de micro-organismos patogênicos, como bactérias multirresistentes, helmintos, protozoários e vírus, além de desinfetantes, hormônios, antibióticos, metais pesados e drogas não metabolizadas por pacientes (Kümmerer, 2001; Emmanuel *et al.*, 2005). A maioria destes patógenos é de origem entérica, sendo, portanto, excretados na matéria fecal, as quais podem contaminar o meio ambiente, tendo acesso novamente a novos hospedeiros através de ingestão da água contaminada. Eles estão presentes em maior concentração nos esgotos hospitalares do que em esgotos provenientes de outras fontes, e a concentração dos antimicrobianos na água que recebe o esgoto hospitalar também é maior, podendo criar um ambiente com forte pressão seletiva (Fuentefria *et al.*, 2008). A falta de um tratamento adequado para a água oriunda destes esgotos e a carência de serviços de saneamento básico pode comprometer a qualidade sanitária dos corpos hídricos, além de contribuir para a disseminação desses patógenos no ambiente, provocando a poluição de ambientes naturais e gerando um desequilíbrio ecológico. Desta forma, é de grande importância avaliar a real prevalência dos genes de resistência aos antibióticos neste meio, visando estimar o risco ambiental bem como para a saúde pública, de um lançamento acidental ou irregular de esgotos hospitalares em ambientes aquáticos sem o devido tratamento. Vale ressaltar que o conceito de resistência utilizada nessa tese está relacionado a uma visão geral representada por qualquer mecanismo que reduz a susceptibilidade aos antibióticos e não à definição utilizada nos ambientes clínicos baseada em concentração mínima inibitória.

## 1.2 Resistência bacteriana a antibióticos

### 1.2.1 Resistoma

Na natureza, micro-organismos como os fungos e as bactérias podem produzir antibióticos como proteção aos ataques de outras bactérias, matando ou inibindo o crescimento das suas competidoras, além de garantir nutrientes para sua sobrevivência, principalmente em ambientes com poucas fontes de nutrientes (Krause *et al.*, 2006). A produção de antibiótico pela bactéria exige que a célula produtora disponha de mecanismos de resistência que a protejam da ação desses compostos. Entretanto, esses mecanismos de resistência podem ser transferidos para outras bactérias, levando aos problemas clínicos de combate às infecções bacterianas como consequência da resistência aos antibióticos administrados. Essa situação vem ameaçando de forma crescente a saúde pública mundial devido ao difícil tratamento de infecções hospitalares causadas pelos principais patógenos (Streit & Daniel, 2010).

O uso excessivo e incorreto de antibióticos na medicina humana e veterinária são os principais desencadeadores do desenvolvimento e da dispersão de bactérias multirresistentes em todo o mundo. As comunidades bacterianas geralmente respondem à seleção de antibióticos através de elementos genéticos móveis que podem ser transferidos horizontalmente entre espécies ou através de mudanças genéticas (mutações) (Kristiansson *et al.*, 2011). A transferência horizontal de genes resistentes a antibióticos pode ocorrer entre cepas bacterianas que não são relacionadas, evolutivamente e ecologicamente, mesmo na ausência de antibióticos (Kruse & Sorum, 1994). Esse mecanismo de transferência é muito importante na aceleração da dispersão desses genes no ambiente. Na maioria das bactérias, principalmente em Gram-negativas, os plasmídeos são os vetores mais importantes desses genes (Hawkey & Jones, 2009). Os transpósons também ajudam a ampliar a mobilidade genética, pois podem transpor alguns genes de resistência desde os plasmídeos até os cromossomos (Williams, 1999). O uso de antibióticos de forma inapropriada vem acelerando a seleção de novas variantes de genes de resistência a antibióticos (GRAs) em microrganismos patogênicos como *Escherichia coli*, *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Klebsiella pneumoniae*, entre outras. Esta resistência pode ser transmitida do ambiente aos humanos, levando risco à saúde humana e de animais (Rodriguez-Brito *et*

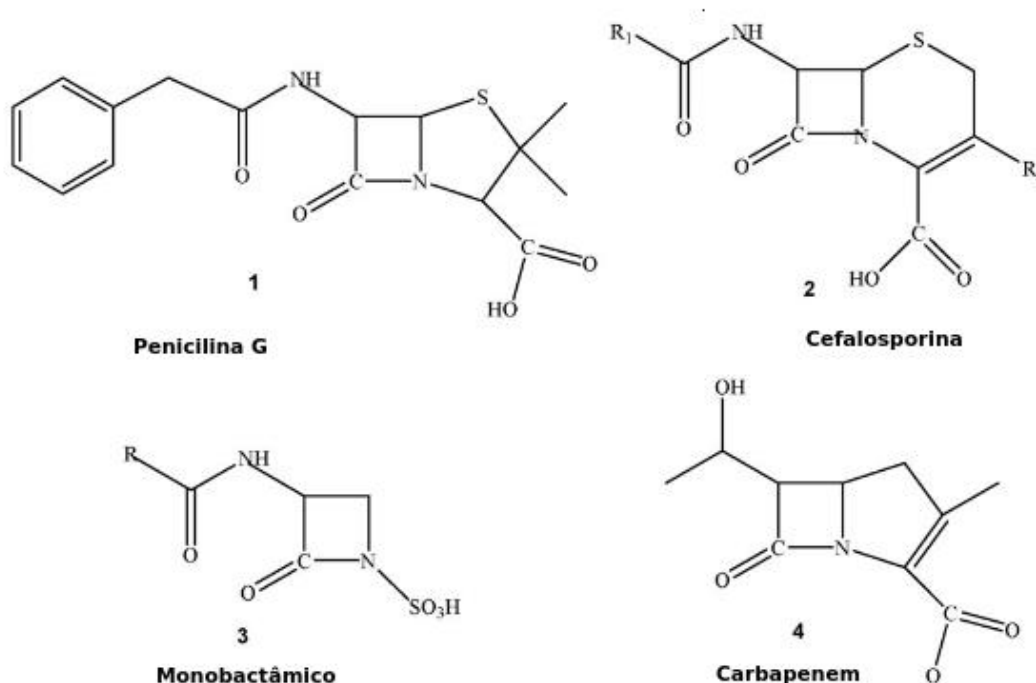
*al.*, 2006). Apesar da grande relevância desse assunto, estudos relacionados à resistência a antibióticos estão geralmente focados em isolados bacterianos obtidos de ambientes clínicos (Henriques *et al.*, 2006). No entanto, os determinantes de resistência a antibióticos encontrados em isolados patógenos clínicos e veterinários compõem apenas uma pequena parcela dos GRAs totais pesquisadas, implicando que o principal reservatório para os GRAs são as bactérias ambientais não patogênicas, particularmente as não cultiváveis (Riesenfeld *et al.*, 2004; Henriques *et al.*, 2006). Esse “pool” de GRAs foi recentemente denominado de “resistoma ambiental de antibióticos” e carrega uma grande diversidade molecular (Wright, 2007). A preocupação do resistoma em relação ao crescente problema da resistência aos antibióticos em patógenos humanos é devido ao provável contato entre patógenos oportunistas como *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Stenotrophomonas maltophilia* e bactérias ambientais não patogênicas (Baquero *et al.*, 2008; Martinez, 2009).

Existem evidências crescentes de que alguns desses genes presentes nos patógenos tenham surgido do resistoma ambiental. Como exemplo temos o CTX-M, uma nova  $\beta$ -lactamase pertencente à classe A, com propriedades de espectro estendido, que parece ter vindo de genes cromossomais da bactéria pertencente ao gênero *Kluyvera* (Wright, 2010). Os GRAs de importância clínica estão presentes em várias bactérias ambientais não patogênicas (Heuer *et al.*, 2002; Riesenfeld *et al.*, 2004; Martinez, 2009) que habitam solos e na água (Baquero *et al.*, 2008). Em vários casos, esses ambientes, que estão gerando populações significativas de GRAs, são poluídos com uma variedade de substâncias químicas, incluindo antibióticos, devido à ação antropogênica (Allen *et al.*, 2010). O contato de bactérias associadas à microbiota humana com micro-organismos ambientais em estações de tratamento ou em ecossistemas naturais é uma característica importante a ser considerada para a compreensão da emergência de novos mecanismos de resistência em patógenos humanos (Baquero *et al.*, 2008).

### **1.2.2 $\beta$ -lactâmicos**

Os antibióticos  $\beta$ -lactâmicos são um importante componente no tratamento de doenças infecciosas causadas por bactérias Gram-negativas e Gram-positivas. Eles representam mais de 65% do mercado de antibióticos

mundial com mais de 50 tipos de fármacos dessa classe comercializados, incluindo as penicilinas, cefalosporinas, carbapenemas, monobactâmicos e, mais recentemente, os híbridos penicilinas-cefalosporinas, e os penems (ex: faropenem). Eles são caracterizados por possuírem um anel  $\beta$ -lactâmico (Bartlett, 2003; Dalhoff & Thomson, 2003) (Figura 1.1).



**Figura 1.1** Estrutura química de antibióticos  $\beta$ -lactâmicos: (1) penicilina, (2) cefalosporina, (3) monobactâmico e (4) carbapenem (Babic *et al.*, 2006).

Todos os fármacos  $\beta$ -lactâmicos são inibidores da síntese da parede celular bacteriana e, portanto, mostram-se ativos contra bactérias em crescimento. A etapa inicial da ação do fármaco consiste na ligação destes receptores celulares (que são as proteínas que se ligam a penicilina ou PBPs-“Penicillin Binding Proteins”). Diferentes receptores possuem afinidades distintas a esses fármacos podendo mediar um efeito distinto de cada um. Além disso, o controle das PBPs é de origem cromossômica e a ocorrência de mutações pode alterar seu número ou sua afinidade pelos  $\beta$ -lactâmicos (Guilfoile, 2007). A integridade do anel é necessária para a atividade, resultando na inativação de um grupo de transpeptidases que catalisam as reações finais de ligação da síntese de peptidoglicano. Eventos de mutação resultam na modificação das PBPs ou na permeabilidade celular, que podem levar a resistência aos  $\beta$ -lactâmicos (Guilfoile, 2007).



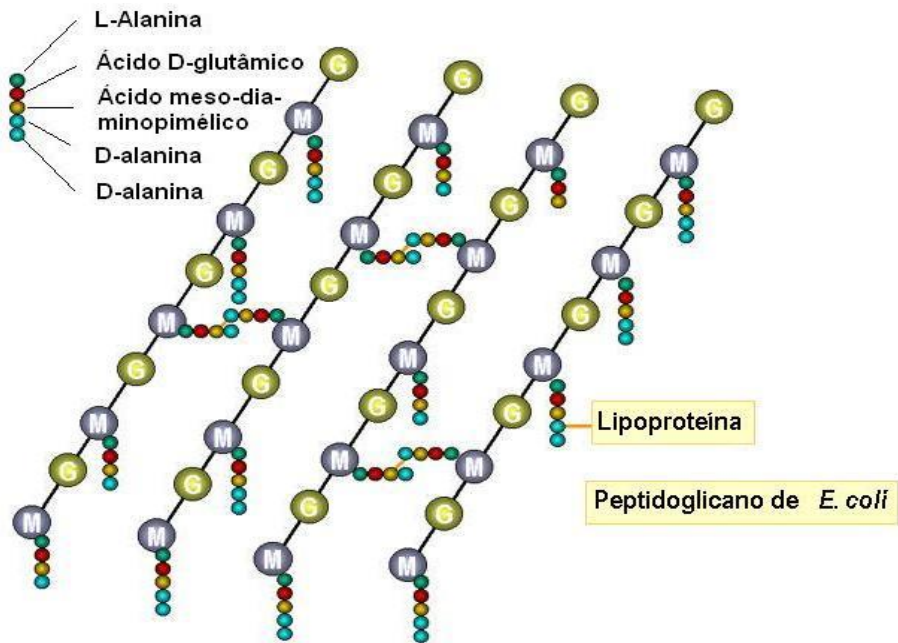
### 1.3 Alvo dos $\beta$ -lactâmicos

Os alvos dos compostos  $\beta$ -lactâmicos são as DD-peptidases (também chamadas por PBPs), as quais são responsáveis pela síntese e remodelagem do peptideoglicano. Essas enzimas possuem serina em seu sítio ativo e realizam seu ciclo catalítico de acordo com mecanismos de acetilação/desacetilação. Os antibióticos  $\beta$ -lactâmicos acilam o sítio ativo das DD-peptidases, formando acil-enzimas covalentes não catalíticas e estáveis, resultando na formação de um peptideoglicano não funcional o que, eventualmente, leva à morte celular (Matagne *et al.*, 1998). Logo, os compostos  $\beta$ -lactâmicos, por mimetizarem o substrato das PBPs, são reconhecidos no lugar dos substratos naturais inativando as DD-peptidases/PBPs. As PBPs são proteínas que contêm o domínio DD-peptidase, mais utilizado para a hidrólise da ligação do D-alanil-D-alanina na fração do pentapeptídeo de peptideglicanos. Vários grupos de enzimas pertencem a essa família como as carboxipeptidases, as transpeptidases e as endopeptidases. As carboxipeptidases removem o terminal D-alanina do pentapeptídeo, as transpeptidases ajudam a formar a ligação cruzada entre os pentapeptídeos das glicanas adjacentes, enquanto que as endopeptidases hidrolisam essas ligações cruzadas. A reação de transpeptidação envolve a perda de uma D-alanina do pentapeptídeo (Ghuysen, 1991).

As PBPs são divididas em dois grupos: as de baixo peso molecular e as de alto peso molecular, as quais por sua vez são subdivididas em três classes com base na similaridade das sequências de aminoácidos (Ghuysen, 1991). As proteínas pertencentes à classe A são bifuncionais e possuem um domínio de glicotransferase. As proteínas da classe B são bifuncionais e também possuem um papel importante na morfogênese da célula através da interação com outras proteínas envolvidas no ciclo celular. Já as proteínas da classe C são menores, monofuncionais, e são divididas em vários tipos (Tipo-4, Tipo-5, Tipo-7 e Tipo AmpH) (Sauvage, 2008). As PBPs compartilham uma atividade comum de DD-peptidase, sendo atividade de DD-transpeptidase, de DD-carboxipeptidase ou de DD-endo-peptidase (Ghuysen, 1991; Goffin & Ghuysen, 1998).

### 1.3.1 Parede celular bacteriana: composição geral

Estima-se que as bactérias surgiram a mais de 3,5 bilhões de anos atrás (Holland, 1997; Manchester *et al.*, 1995). Uma característica estrutural primária da bactéria é sua parede celular cuja função é indispensável para o organismo conferindo suporte na manutenção da morfologia bacteriana (Massova & Mobashery, 1998). As bactérias possuem uma camada de três fitas de peptidoglicano, responsável pela forma e rigidez das células, protegendo-a de um possível rompimento devido à pressão osmótica entre os meios externo e interno. O peptidoglicano é composto pela repetição alternada de dois açúcares aminados, o ácido N-acetilglicosamina e o ácido N-acetilmurâmico, além de um tetrapeptídeo, sempre ligado ao resíduo de N-acetilmurâmico. O ácido N-acetilmurâmico e a N-acetilglucosamina estão ligados entre si por uma ligação glicosídica ( $\beta 1 \rightarrow 4$ ) (Guilfoile, 2007). Essa composição varia entre bactérias Gram-negativas e Gram-positivas, mas sempre termina na formação de um dipeptídeo D-alanina-D-alanina. O dipeptídeo sofre uma adição de mais três aminoácidos e um amino açúcar, o ácido murâmico, gerando um muramil pentapeptídeo. Esse pentapeptídeo-açúcar se liga a uma molécula de acetilglicosamina, tornando-se o peptidoglicano propriamente dito. As unidades de peptidoglicano são produzidas dentro da célula bacteriana, mas a ligação cruzada é catalisada na parte externa da membrana citoplasmática, por um grupo de enzimas ancoradas à membrana, as transpeptidases da parede celular. Nessa reação cruzada, uma ligação peptídica é formada entre a penúltima D-alanina de uma cadeia e o resíduo de aminoácido livre do ácido diaminopimélico (em Gram-negativas) ou uma L-lisina (em Gram-positivas) da outra cadeia (Wilke *et al.*, 2005). A Figura 1.2 ilustra a estrutura da molécula peptidoglicano de uma bactéria Gram-negativa.

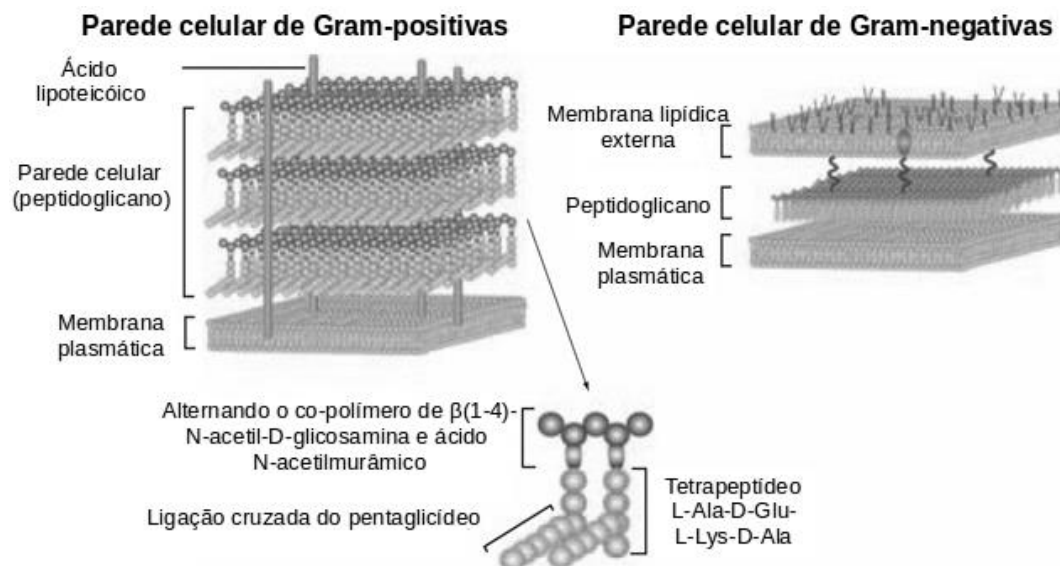


**Figura 1.2. Esquema da composição da molécula de peptidoglicano de uma bactéria Gram-negativa, a *E. coli*. M: N-acetil murâmico, G: N-acetil glicosamina.**

*Fonte:* [pathmicro.med.sc.edu/portuguese/chapter\\_4\\_bp.htm](http://pathmicro.med.sc.edu/portuguese/chapter_4_bp.htm)

### 1.3.2 Parede celular bacteriana: Gram-positivas e Gram-negativas

Em bactérias Gram-positivas a parede celular é composta por uma camada espessa de peptidoglicano, ligado ao ácido teicóico (seus esqueletos geralmente são polímeros fosforados de ribitol ou glicerol) ou ácido teicurônico (ácido glicurônico contendo polissacarídeos). Em Gram-negativas a parede celular bacteriana é composta por uma camada de peptidoglicano e por mais três componentes que a envolvem externamente, que são: lipoproteína, membrana externa e lipopolissacarídeo. O peptidoglicano localiza-se no espaço periplasmático, entre a membrana interna e a externa, que consiste em uma dupla camada lipídica (Guilfoile, 2007). A Figura 1.3 ilustra o esquema de uma parede celular de bactérias Gram-positivas e bactérias Gram-negativas.



**Figura 1.3** Esquema da composição da parede celular de bactérias Gram-positivas e Gram-negativas e a disposição de seus componentes.

Fonte: [http:// www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/peptidoglycans.html](http://www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/peptidoglycans.html)

#### 1.4 $\beta$ -lactamases

A resistência bacteriana aos antibióticos  $\beta$ -lactâmicos pode ser obtida através de três estratégias, basicamente: (i) produção de enzimas que hidrolisam os  $\beta$ -lactâmicos, (ii) prevenção da interação do fármaco com o alvo, e (iii) expulsão dos antibióticos através de bombas de efluxo, por bactérias Gram-negativas, impedindo o acesso destes ao seu alvo. O mecanismo mais comum é a produção de enzimas que degradam ou modificam o antibiótico antes que ele atinja o sítio alvo (Wright, 2005; Wilke *et al.*, 2005).

Dentre as enzimas que conferem resistência pela inativação de antibióticos por hidrólise, destacam-se as  $\beta$ -lactamases (E.C.3.5.2.6), as macrolídeos esterases e as epoxidases (Wright, 2005). Em bactérias Gram-positivas as  $\beta$ -lactamases produzidas são liberadas na área extracelular e aí se ligam ao  $\beta$ -lactâmico, clivando-o. Já nas bactérias Gram-negativas, as  $\beta$ -lactamases são liberadas no espaço periplasmático, tornando esse mecanismo de clivagem e inibição do antibiótico mais efetivo (Ghuysen, 1991). Em bactérias Gram-negativas as  $\beta$ -lactamases representam o mecanismo mais importante de resistência a antibióticos. Elas são produzidas em resposta ao uso de  $\beta$ -lactâmicos e são responsáveis pela quebra da ligação de amina no anel  $\beta$ -lactâmico do antibiótico tornando-o inócuo (Petrosino *et al.*, 1998). Os mecanismos de resistência aos  $\beta$ -lactâmicos incluem inacessibilidade do

antibiótico à enzima alvo, modificações das enzimas alvo e/ou inativação direta dos antibióticos pelas  $\beta$ -lactamases (Zhang *et al.*, 2009). Esse grupo de enzimas possui diversas variantes com diferentes espectros de atividade, sendo que novas formas são frequentemente descritas por todo o mundo, inclusive as mais preocupantes, as  $\beta$ -lactamases de espectro estendido (ESBLs – “Extended-spectrum  $\beta$ -lactamases”). As ESBLs são capazes de hidrolisar as cefalosporinas de amplo espectro, como a cefotaxima e ceftazidima, e os monobactâmicos, como o aztreonam (Jacoby & Medeiros, 1991) e pertencem às famílias TEM, SHV, CTX-M (classe A) e OXA (classe D). Estas enzimas são codificadas por genes que evoluem rapidamente e, portanto, constituem um grupo caracterizado por alta variabilidade. Os genes que codificam as  $\beta$ -lactamases estão presentes em ambas as bactérias Gram-positivas e Gram-negativas, podendo ser encontrados em plasmídeos, transpósons, cromossomos ou integrons (Petrosino *et al.*, 1998; Henriques *et al.*, 2006).

A similaridade entre as  $\beta$ -lactamases e DD-peptidases não está apenas restrita às propriedades “mecanísticas”. Estudos envolvendo difração por raios-X mostraram analogias estruturais entre DD-peptidase de *Streptomyces* R61 e várias  $\beta$ -lactamases pertencentes às classes A e C (Kelly *et al.*, 1986; Saramou *et al.*, 1986; Oefner *et al.*, 1990; Lobkovsky *et al.*, 1993; Jamin *et al.*, 1995). Além disso, o domínio responsável pela atividade de transpeptidase, das DD-peptidases possuem uma similaridade marcante com o sítio ativo das  $\beta$ -lactamases. Sendo assim, esses dados estruturais, assim como de outros trabalhos (Lobkovsky, 1994; Maveyraud *et al.*, 1996; Bulychev *et al.*, 1997), forneceram evidências que suportam a existência de uma relação próxima entre essas duas famílias de enzimas que reconhecem penicilinas e a hipótese de que essas duas famílias compartilham um ancestral comum.

#### **1.4.1 Classificação das $\beta$ -lactamases**

Existem dois sistemas de classificação envolvendo as  $\beta$ -lactamases. O primeiro foi criado por Bush (Bush, 1989; Bush *et al.*, 1995) e baseia-se nas características funcionais. De acordo com essa classificação as  $\beta$ -lactamases são divididas em quatro grupos definidos de acordo com os seus substratos e sensibilidade ao inibidor ácido clavulânico e EDTA. O grupo 1 é composto pelas  $\beta$ -lactamases que não são inibidas pelo ácido clavulânico (pertencem a

classe C de Ambler); o grupo 2 contém as  $\beta$ -lactamases inibidas pelo ácido clavulânico (pertencem às classes A e D de Ambler); o grupo 3 inclui as enzimas que possuem o íon zinco em seu sítio ativo e não são inibidas pelo ácido clavulânico mas são pelo EDTA (pertencem à classe B de Ambler); finalmente, o grupo 4 é composto pelas  $\beta$ -lactamases que não são inibidas pelo ácido clavulânico mas não se encaixam nos outros grupos (Bush, 1995).

A segunda forma de classificação foi sugerida por Ambler (1980) com base na similaridade de sequências de aminoácidos. A partir daí essas enzimas foram separadas em quatro grandes classes moleculares (A, B, C e D). As classes A, C e D compartilham dobramentos similares e constituem evolutivamente grupos distintos de serina  $\beta$ -lactamases, enquanto a classe B são metaloenzimas e parecem ser análogas, diferindo completamente das serina  $\beta$ -lactamases em relação às sequências, ao seu enovelamento e ao mecanismo enzimático. Essa classe é caracterizada por conter íons de zinco para romper o anel  $\beta$ -lactâmico (Livermore, 1995; Wilke *et al.*, 2005). Por esta razão as classes A, C e D foram definidas como sendo “Serina  $\beta$ -lactamases” e a classe B como sendo as “Metallo  $\beta$ -lactamases”. Bactérias que possuem genes que codificam metallo  $\beta$ -lactamases são consideradas por terem fenótipos mais resistentes encontrados na área clínica. As  $\beta$ -lactamases pertencentes à classe A são conhecidas como penicilinases por conseguir hidrolisar muito bem a penicilina e ampicilina, e até algumas cefalosporinas, a classe B carbapenêmicos e outros  $\beta$ -lactâmicos, a classe C por hidrolisarem uma ampla gama de cefalosporinas e a classe D é conhecida como oxacilinases por serem especialistas na hidrólise de oxacilinas, cloxacilina, meticilina e carbenicilina (Hall & Barlow, 2004). Além desses  $\beta$ -lactâmicos, as enzimas da classe D, assim com as da classe A, se especializaram na hidrólise de grupos específicos de cefalosporinas (Naas & Nordmann, 1999). Exemplos de bactérias que produzem essas enzimas são as pertencentes à família Enterobacteriaceae e *Haemophilus influenza*, *Neisseria gonorrhoeae*, *Vibrio cholerae*, *Pseudomonas aeruginosa* e *Moraxella* spp., (Garau, 1994). Além disso, são também detectadas em bactérias anaeróbias como espécies do gênero *Bacteroides* (como *B. fragilis*), *Prevotella*, *Porphyromonas*, *Bilophila wadsworthia*, *Fusobacterium* spp. e *Clostridium* spp. (Summanen *et al.*, 1993; Könönen *et al.*, 1998).

### 1.4.2 Mecanismo de atividade

O resíduo serina do sítio ativo da  $\beta$ -lactamase ataca a ligação do  $\beta$ -lactâmico, formando uma acil-enzima intermediária e covalente. A enzima intermediária acil é então hidrolisada por uma molécula de água, inativando o antibiótico (Goldberg *et al.*, 2003)(Figura 1.4).



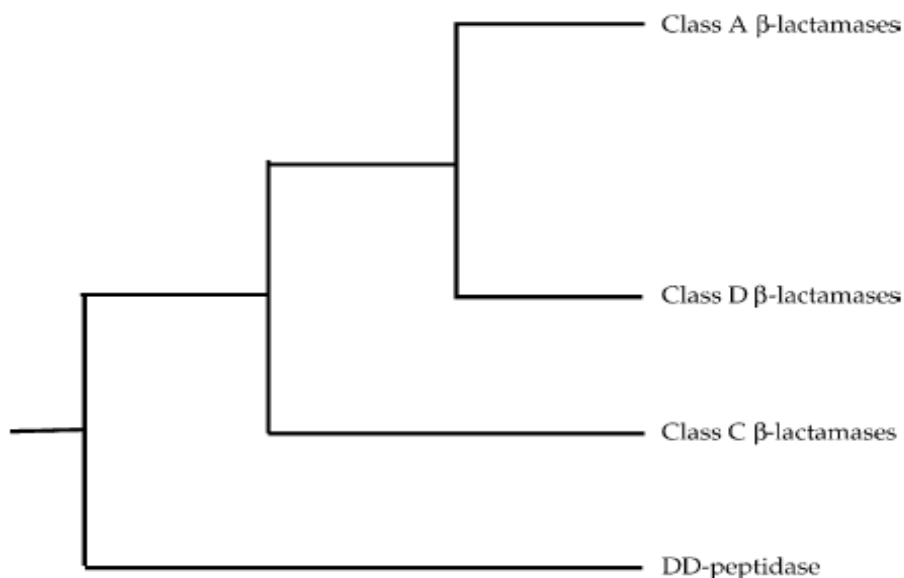
**Figura 1.4. Reação enzimática das  $\beta$ -lactamases (3.2.5.8).** Fonte: MetaCyc – [www.biocyc.org](http://www.biocyc.org)

O mecanismo catalítico das serina  $\beta$ -lactamases envolve uma etapa de acilação que é comum tanto em  $\beta$ -lactamases quanto nas PBPs. Já a outra etapa, de desacilação só é vista em serina  $\beta$ -lactamases (Hata *et al.*, 2006). Logo, a essência da resistência aos  $\beta$ -lactâmicos pelas serina- $\beta$ -lactamases surge da reação de desacilação. A fase de acilação ocorre quando a ligação peptídica é quebrada e uma ligação éster é formada entre o carbono da carbonila e a enzima. Já a fase de desacilação se dá quando a ligação éster é hidrolisada e a enzima não acilada é regenerada (Hata *et al.*, 2006).

### 1.4.3 Filogenia e características moleculares

A filogenia dos genes de resistência a antibióticos é bastante complexa uma vez que estes estão presentes em muitos elementos genéticos móveis, sendo transferidos de forma horizontal e, desta forma, estão presentes em uma ampla gama de hospedeiros (Hall & Barlow, 2004). Sendo assim é difícil estabelecer uma correlação confiável entre a origem do gene e o organismo no qual esse gene foi encontrado. Ambos os grupos de  $\beta$ -lactamases (serina e metalo) são muito antigos e as classes pertencentes a eles são muito diversificadas sendo que diversos traços de similaridade entre elas, em relação às sequências de aminoácidos, foram perdidas (Hall & Barlow, 2004). Mesmo assim, a filogenia baseada na estrutura das proteínas foi capaz de reconstruir a evolução de ambos os grupos de  $\beta$ -lactamases e inferir que essas enzimas mais antigas foram originadas a mais de dois bilhões de anos atrás, com

algumas serina  $\beta$ -lactamases presentes em plasmídeos por milhões de anos, bem antes do início do uso de antibióticos.



**Figura 1.5** Árvore filogenética das três classes de serina  $\beta$ -lactamases, a partir do alinhamento estrutural, junto com seu homólogo, as DD-peptidases (Hall & Barlow, 2003; Garau *et al.*, 2004).

Hall e Barlow (2003) construíram uma árvore filogenética das serina  $\beta$ -lactamases utilizando métodos originalmente desenvolvidos para alinhamento de sequências de proteínas, para construir alinhamentos baseados em estrutura de proteínas (alinhamento estrutural). A metodologia de alinhamento baseado em estrutura de proteínas permitiu identificar aminoácidos homólogos baseados em suas posições dentro da estrutura proteica. Os resultados desse trabalho revelaram que as  $\beta$ -lactamases pertencentes à classe C estão mais próximas evolutivamente das PBP/DD-peptidases do que as classes A e D e sugeriram que as classes A e D tenham evoluído a partir de uma forma ancestral da classe C (Figura 1.5).

O primeiro antibiótico a ser descoberto foi a penicilina, por Alexander Fleming, em 1928, um  $\beta$ -lactâmico. Com o uso mundial de antibióticos, especialmente os  $\beta$ -lactâmicos, vários derivados químicos foram produzidos a partir da penicilina para combater a resistência que as bactérias desenvolveram. Esses derivados químicos são chamados de  $\beta$ -lactâmicos de espectro estendido (ESBL). Não existe um consenso na definição precisa das ESBLs, mas a definição mais comum utilizada é que são  $\beta$ -lactamases capazes de conferir resistência bacteriana à penicilina, terceira e quarta gerações de



cefalosporinas, e aztreonam (mas não às cefamicinas e carbapenemas), pela hidrólise dos antibióticos, e são inibidas por inibidores de  $\beta$ -lactamases como ácido clavulânico, sulbactam e tazobactam (Paterson & Bonomo, 2005; Silva & Lincopan, 2012). Porém, as bactérias continuaram desenvolvendo resistência a esses novos fármacos. As enzimas capazes de clivar esses fármacos são conhecidas como  $\beta$ -lactamases de espectro estendido (ESBLs) e foram primeiras observadas no início dos anos 80 (Paterson & Bonomo, 2005). As ESBLs mais comuns incluem as  $\beta$ -lactamases das famílias: TEM, SHV e CTX-M (classe A), e as  $\beta$ -lactamases classe D da família OXA. As ESBLs se desenvolveram a partir da difusão das famílias de enzimas codificadas por plasmídeos como TEM (classe A), SHV (classe A), CTX-M (classe A) e OXA (classe D). TEM-1 é capaz de hidrolisar penicilinas e cefalosporinas de primeira geração, mas não consegue atacar a cefalosporina oximino (Gupta, 2007). As ESBLs derivadas de TEM-1, TEM-2 e SHV-1 diferem de seus progenitores por apenas um aminoácido. Porém, isso resulta em uma profunda mudança na atividade enzimática das ESBLs, permitindo-as hidrolisar a terceira geração de cefalosporinas e aztreonam (Paterson & Bonomo, 2005). Até o presente foram identificados 205 tipos de TEM, 168 tipos de SHV, 148 tipos de CTX-M e 18 tipos de OXA, de espectro estendido (<http://www.lahey.org/Studies>) sendo 61 ESBL em TEM-1, 12 na família SHV e mais de 80 na família CTX-M (Novais *et al.*, 2010).

#### **1.4.4 Especificações sobre cada classe de serina- $\beta$ -lactamase**

**Classe A:** foram as primeiras  $\beta$ -lactamases a serem identificadas em isolados de bactérias resistentes a penicilina e continuam sendo as mais abundantes até hoje. Por volta de 1940, poucos anos após a introdução da penicilina na área clínica, foi detectada a resistência à penicilina em *Staphylococcus aureus*, a partir da produção de  $\beta$ -lactamase classe A (Medeiros, 1997). Em 1963 foi descoberto o TEM-1, uma  $\beta$ -lactamase classe A plasmidial, a mais comum atualmente (Bradford, 2001). Ela é considerada uma enzima totalmente eficiente uma vez que hidrolisa a penicilina e seus derivados (como ampicilina, carbenicilina e piperaciclina). A família SHV, relacionada ao TEM (possui 68% de similaridade de aminoácidos com TEM-1), foi descoberta em 1974 (Matthew *et al.*, 1979). O SHV-1 é o tipo mais comum, hoje, encontrado em plasmídeos (Leung *et al.*, 1997). Ambos são bem distribuídos

ao longo do mundo. Elas são encontradas em várias enterobactérias, incluindo *Salmonella typhimurium*, *E. coli*, *Klebsiella pneumoniae* e *Proteus mirabilis* (Tzouvelekis *et al.*, 2000). As enzimas que compõem a classe A compartilham uma similaridade de aminoácidos acima de 25% (Sanschagrín *et al.*, 1995).

**Classe C:** A classe C inclui as AmpC que contribuem para resistência a múltiplos  $\beta$ -lactâmicos e são produzidas, até o momento, apenas por patógenos Gram-negativos, particularmente os da família Enterobacteriaceae, e podem estar presentes em cromossomos ou plasmídeos. Baseado na similaridade de sequência de aminoácidos, as AmpC podem ser divididas em cinco famílias, C-1 até C-5 (Walther-Rasmussen & Hoiby, 2002). Elas não são ESBLs por serem resistentes aos inibidores clavulanato, sulbactam e tazobactam (Silva & Lincopan, 2012). Além disso, as AmpC parecem ter uma função adicional como uma enzima que hidrolisa peptidoglicano (Bishop & Weiner, 1993). Outras famílias como MOX, DHA-1 e ACT-1 também fazem parte dessa classe (Majiduddin *et al.*, 2002).

**Classe B:** O grupo das metalo  $\beta$ -lactamases compõe a classe B e é classificado, até agora, em três subclasses: B1, B2 e B3, baseado em fatores funcionais e mecânicos. Entretanto, compartilham um baixo grau de similaridade entre si (< 20%). Dessa forma, o alinhamento clássico de aminoácidos desta classe não mostra resultados fidedignos (Garau *et al.*, 2004).

**Classe D:** também conhecidas como OXA  $\beta$ -lactamases (OXAs) constituem a classe de  $\beta$ -lactamases mais diversa dentre as quatro classes moleculares. Essa diversidade pode ser observada nos níveis genéticos e bioquímicos, com enzimas possuindo tanto um espectro de atividade estreito quanto um estendido, resultantes de mutações pontuais. A maioria das enzimas pertencentes à classe D encontram-se em integrons de classe 1, mas também há relatos da presença desta flanqueada por sequências de inserção e associadas a transpósons (Naas & Nordmann, 1999). A similaridade entre as sequências de DNA das classes A e C são limitadas, sendo basicamente restritas às três regiões principais em volta do sítio ativo (Majiduddin *et al.*, 2002). De uma forma geral, as enzimas da classe D apresentam aproximadamente 16% de similaridade com as enzimas da classe A ou classe C (Naas & Nordmann, 1999). As enzimas da classe D tipo carbapenemases podem ser classificadas em oito subgrupos, sendo a similaridade entre

sequências de aminoácidos do mesmo subgrupo maior que 92% enquanto que a similaridade entre sequências de aminoácidos de diferentes subgrupos varia entre 40-70% (Rasmussen & Høiby, 2007).

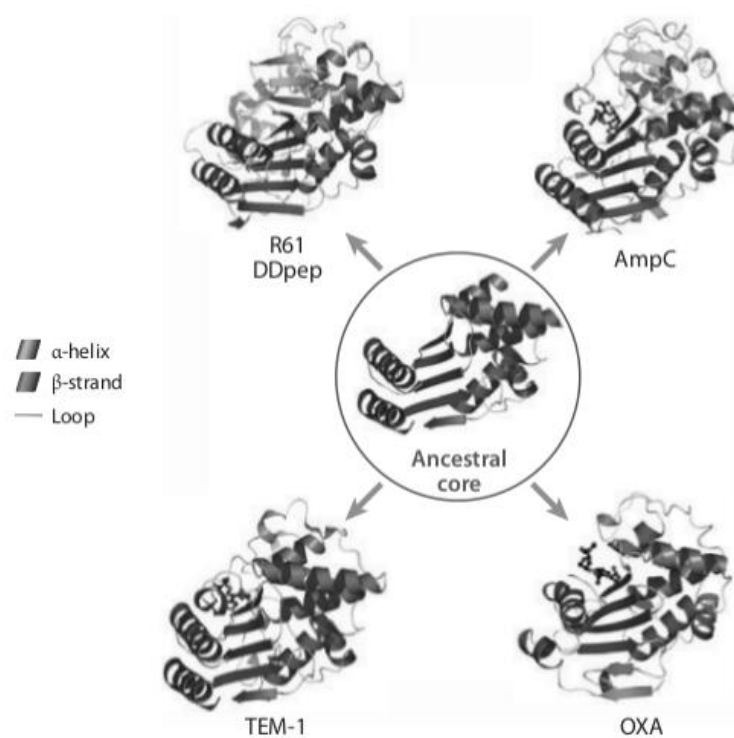
#### **1.4.5 Sítio ativo das PBPs e serina- $\beta$ -lactamases**

Na vizinhança iminente ao resíduo de serina, localizada no sítio ativo das DD-peptidases e  $\beta$ -lactamases, vários elementos conservados que parecem estar direta ou indiretamente envolvidos no reconhecimento do substrato e no processo catalítico já foram identificados (Matagne *et al.*, 1998). O primeiro elemento contém a serina no sítio ativo e, próximo a ele, um resíduo de lisina, com dois aminoácidos variáveis no meio (entre a serina e lisina) (S-Xaa-Xaa-K). As  $\beta$ -lactamases e PBPs compartilham várias regiões altamente conservadas. Todas as enzimas da superfamília das enzimas que se ligam a penicilina ( $\beta$ -lactamases e PBPs) possuem uma tétrade de aminoácidos em comum: S-X-X-K, onde o X representa qualquer aminoácido. Essa região compõe o sítio catalítico. Além disso, todas as enzimas que se ligam a penicilinas também possuem uma tríade Lys (His)-Thr (Ser)-Gly [K(H)-T(S)-G], que forma parte da folha beta pregueada que delimita uma parede da cavidade de ligação. Outra tríade em comum é a Ser (Tyr)-Asp-Asn [S(Y)-D-N] (Medeiros, 1997). As enzimas pertencentes à classe A possuem três regiões conservadas: o motivo Ser-x-x-Lys [S-X-X-K], Ser-Asp-Asn [S-D-N] e Lys-Thr-Gly [K-T-G], nas posições 70, 130 e 234, respectivamente (Singh *et al.*, 2009). A região Arg-x-Glu-x-x-Leu-Asn (Ser) [R-X-E-X-X-L-N(S)], também é altamente conservada em  $\beta$ -lactamases da classe A e são alteradas nas outras enzimas que reconhecem penicilinas (Medeiros, 1997). Já as pertencentes à classe C possuem motivos específicos como: S-X-S-K, Y-S(A)-N e K-T-G, nas posições 64, 150 e 314, respectivamente (Singh *et al.*, 2009). As enzimas da classe D possuem basicamente entre si três elementos do sítio ativo altamente conservados. O primeiro elemento é a tétrade, S-X-X-K, o segundo elemento, S-X-V(I), são análogos ao motivo S-D-N na classe A, e Y-S(A)-N nas  $\beta$ -lactamases pertencentes à classe C. Outros motivos conservados são a tríade Y(F)-G-N e a tétrade Trp-x-x-Gly [W-X-X-G] que não possuem análogos nas classes A e C e o motivo K-T-G comum às três classes de  $\beta$ -lactamases (Walther-Rasmussen & Høiby, 2007; Singh *et al.*, 2009) (Tabela 1.6). A Figura 1.6 ilustra exemplos de estruturas terciárias das classes A, C e D das  $\beta$ -

lactamases e sua homóloga DD-peptidase, assim como o “core” de seu ancestral.

**Figura 1.6. Estrutura molecular do ancestral das serina  $\beta$ -lactamases, da enzima homóloga DD-peptidase R61 e das  $\beta$ -lactamases classe A (TEM), classe C (AmpC) e classe D (OXA) (Fonte: Morar & Wright, 2010).**

Apesar de essas regiões serem altamente conservadas, já foram descritas substituições de alguns aminoácidos dessas regiões em  $\beta$ -lactamases (Matagne *et al.*, 1998), como apresentado na Tabela 1.1.



**Tabela 1.1 Principais regiões conservadas (motivos) referentes às classes de serina  $\beta$ -lactamases. Ser: Serina, Xaa: um aminoácido qualquer, Lys: Lisina, Asp: Ácido aspártico, Asn: Asparagina, Gly: glicina, Arg: Arginina, Thr: Treonina, Ala: Alanina, Tyr: Tirosina, His: Histidina.**

Classe	Motivo 1	Motivo 2	Motivo 3
A	<sup>70</sup> Ser-Xaa-Xaa-Lys	<sup>130</sup> Ser-Asp-Asn	<sup>234</sup> Lys-Thr-Gly
		Ser-Asp-Ser	Lys-Ser-Gly
		Ser-Asp-Gly	Arg-Thr-Gly Arg-Ser-Gly
C	<sup>64</sup> Ser-Xaa-Ser-Lys	<sup>130</sup> Tyr-Ala-Asn	<sup>314</sup> Lys-Thr-Gly
		Tyr-Ser-Asn	
D	<sup>70</sup> Ser-Xaa-Xaa-Lys	<sup>144</sup> Tyr-Gly-Asn	<sup>298</sup> His-Thr-Gly
			Lys-Thr-Gly
			Lys-Ser-Gly

Fonte: Matagne *et al.*, 1998

A tríade conservada Ser-Asp-Asn [S-D-N] forma um “loop” na estrutura terciária das  $\beta$ -lactamases, formando uma parede da cavidade do sítio catalítico (Jacob *et al.*, 1990). A partir de trabalhos de mutação sítio dirigida (Jacob *et al.*, 1990; Dubus *et al.*, 1994), verificou-se que a substituição da Ser<sup>130</sup> dessa tríade, em  $\beta$ -lactamases da classe A, pode chegar a inativar a enzima ou diminuir sua estabilidade, dependendo do aminoácido substituído. Com isso, a Ser<sup>130</sup> deve estar envolvida na manutenção da estrutura da cavidade do sítio ativo. A substituição do Asp<sup>131</sup> em Glu e Gly mostrou ser altamente prejudicial para a estabilidade da enzima, refletindo em perturbações estruturais. A substituição do Asn<sup>132</sup> em Ala mostrou diminuir drasticamente a reação da atividade enzimática (em mais de 100x) (Jacob *et al.*, 1990). Em relação à classe D os aminoácidos Y-G da tríade Y-G-N parecem ter um papel de catálise através da hidrólise da molécula de água (como o E-166 na classe A) (Sanschagrín *et al.*, 1995). Em relação à tríade Lys-Thr(Ser)-Gly [K-T(S)-G], o resíduo Lys<sup>234</sup> mostrou ser importante na ligação do substrato a enzima e na estabilização desta (Ellerby *et al.*, 1990).

## 1.5 Reconhecimento da função enzimática por similaridade de sequências

A conservação da função enzimática tem sido utilizada para a obtenção do ponto de corte necessário de identidade de sequências para a transferência de função de uma proteína de função conhecida para outra proteína de função desconhecida. Com o desenvolvimento de algoritmos sensíveis como FASTA (Pearson, 2000), BLAST (Altschul *et al.*, 1996), PSI-BLAST (Altschul *et al.*, 1997) e Modelos Ocultos de Cadeia de Markov (HMM) (Eddy, 1998) foi possível aprimorar o reconhecimento de proteínas relacionadas evolutivamente (homólogos). Apesar disso, a inferência funcional a partir de homologia continua propensa a erros e é apenas confiável quando o grau de similaridade é bem alto (Rost *et al.*, 2002). A partir do desenvolvimento contínuo dessas metodologias a habilidade de reconhecimento de homólogos distantes foi bastante aprimorada. Apesar disso, como só deve existir em torno de 1000 principais superfamílias na natureza, a maioria das proteínas homólogas deve ter diferentes funções, dificultando a inferência funcional a partir da similaridade de sequências (Tian *et al.*, 2004).

As três abordagens básicas para a inferência de uma função enzimática a partir das sequências primárias se baseiam em: transferência de função por similaridade, presença de um motivo ou padrão e identificação de resíduos funcionais. A abordagem da transferência de função através da similaridade é a mais utilizada para anotação funcional e envolve duas etapas: a detecção de homólogos e a inferência funcional a partir da homologia (Tian *et al.*, 2004). Geralmente, duas proteínas que são estruturalmente similares entre elas são consideradas relacionadas evolutivamente. Segundo os trabalhos de Rost (1999), Yang & Honig (2000) e Lindahl & Elofsson (2000) foi observado que, quando a identidade entre duas sequências é superior a 30%, elas possuem estruturas similares e são relacionadas evolutivamente. Com base em experimentos do trabalho de Tian & Skolnick (2003) sugere-se que, para anotação funcional, 40% de similaridade de sequências pode ser usada como um corte confiável para a transferência dos três primeiros dígitos de um número EC (“Enzyme Commission”) de uma enzima. Entretanto, para transferir os quatro dígitos de um número EC é necessária uma similaridade acima de 60% entre as sequências de aminoácidos para obter uma acurácia acima de

90%. Devos e Valencia (2000) fizeram um trabalho no qual utilizaram alinhamentos estruturais de FSSP (famílias de proteínas com estruturas similares) e descobriram que, acima de 50% de similaridade entre as sequências, todos os quatro dígitos do número E.C. mantiveram-se conservados.

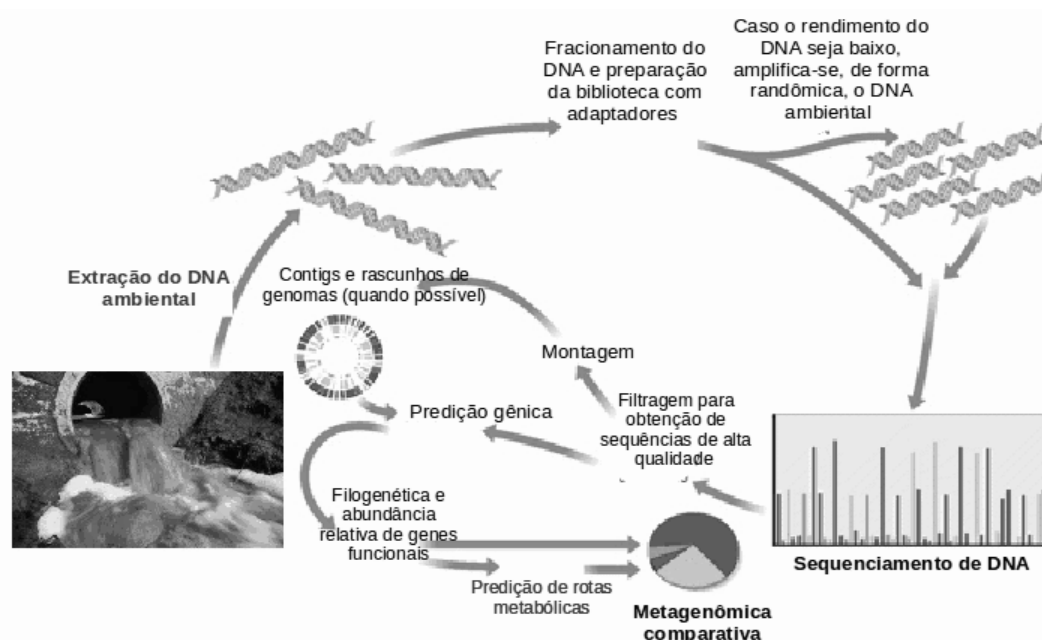
Modelos Ocultos de Markov (HMM) são modelos estatísticos que utilizam alinhamentos múltiplos de sequências homólogas para quantificar as frequências de aminoácidos e as probabilidades de posições específicas de deleções e inserções ao longo do alinhamento (Eddy, 1998). Eles são utilizados na construção de perfis de motivos conservados e famílias de proteínas uma vez que eles possuem mais informações sobre a família da sequência do que uma sequência apenas (Krogh *et al.*, 1994; Eddy, 1998). Os pHMMs (perfis de Cadeia Oculta de Markov) têm sido descritos como muito eficientes na detecção de padrões conservados em um grupo de sequências (Karplus *et al.*, 1997; Eddy *et al.*, 1995; Karchin & Hughey, 1998) e ter uma melhor performance do que métodos que utilizam perfis com apenas uma sequência como o PSI-BLAST (Eddy, 1998, Karplus *et al.*, 2001). A elevada sensibilidade detectada pelos pHMMs faz com que essa técnica seja muito promissora na busca de homólogos distantes com um sucesso considerável (Baldi *et al.*, 1994; Karplus *et al.*, 1997; Srivastava *et al.*, 2007), incluindo em trabalhos envolvendo estudo de Metallo  $\beta$ -lactamases (Widmann *et al.*, 2012).

## 1.6 Metagenômica

No meio ambiente, apenas uma pequena porcentagem de espécies de micro-organismos pode ser identificada por técnicas tradicionais e dependentes de cultivo. Estima-se que aproximadamente 99% dos micro-organismos (especialmente os encontrados no solo) não são cultiváveis devido à dificuldade de identificação das condições necessárias para o cultivo destes ou a necessidade dos mesmos em crescer através de um consórcio microbiano (grupo de diferentes espécies de um micro-organismo que atuam em conjunto, como uma comunidade). Estudos recentes de micro-organismos que ainda não são cultiváveis têm sido realizados através da metagenômica, uma abordagem de análise do DNA ambiental independente de cultivo, aliada aos importantes avanços da bioinformática e das novas tecnologias de sequenciamento (Jones

*et al.*, 2010). O termo metagenômica se originou no final dos anos 90 (Handelsman *et al.*, 1998), mas outros termos como bibliotecas de DNA ambiental (Stein *et al.*, 1996), bibliotecas de DNA de solo (MacNeil *et al.*, 2001), comunidade genômica (Tyson *et al.*, 2004) são encontrados na literatura para descrever a mesma abordagem. Com o surgimento de tecnologias que possibilitam a obtenção de sequências com comprimentos maiores, geradas por sequenciadores de alta vazão, a utilização dessa abordagem tem crescido consideravelmente nos últimos tempos.

A metagenômica consiste na extração e sequenciamento direto do DNA ambiental. A partir daí pode-se utilizar duas abordagens posteriores. Uma delas é a Triagem baseada em sequência (“Sequence-Based Screening”) na qual se busca um gene, um fragmento genômico ou até um genoma completo através do sequenciamento direto ou homologia entre sequências, como por amplificação por PCR (Vergin *et al.*, 1998) ou hibridização (Stein *et al.*, 1996), ou então através de similaridade de domínios conservados (ou regiões) com sequências conhecidas depositadas em bancos públicos curados (Yun *et al.* 2004). A triagem baseada em sequências pode identificar sequências que não seriam expressas pela espécie hospedeira que carrega a biblioteca (Figura 1.7) (Riesenfeld *et al.*, 2004).



**Figura 1.7. Esquema das etapas utilizadas para o estudo da metagenômica, baseada em sequências, a partir de uma amostra ambiental de interesse (Adaptado de Jansson, 2011).**



Nos últimos oito anos houve uma explosão de interesse e atividade na metagenômica, acompanhada por avanços na tecnologia que facilitou estudos em uma escala que antes não era possível quando essa área foi iniciada. Como exemplo, em 2004, o pesquisador Craig Venter e colaboradores (Venter *et al.*, 2004) conduziram um grande projeto de sequenciamento massivo com foco na microbiota do Mar de Sargasso, considerado um ambiente pobre em nutrientes. Esse projeto resultou em dois milhões de sequências com rendimento de mais de 1,6 bilhões de pares de bases de informação, incluindo aproximadamente um bilhão de pares de bases de sequências não redundantes. O sequenciamento desse projeto foi iniciado utilizando-se o método de Sanger (Sanger *et al.*, 1977), mas, pouco tempo depois, em 2005, foi lançado o primeiro sequenciador de alta vazão, o 454/Life Science, que contribuiu para os avanços na área da tecnologia de sequenciamento, e expandiu as abordagens que podem ser consideradas na metagenômica. A área de metagenômica tem mostrado sucesso em diferentes aplicações como no estudo de genes isolados (ex: celulases, Wang *et al.*, 2009; genes de resistência a antibióticos, Forsberg *et al.*, 2012), de rotas metabólicas (ex: síntese de antibióticos, Gillespie *et al.*, 2002), de organismos (ex: Archea, Cavicchioli *et al.*, 2007) e de comunidades (ex: comunidade microbiana de drenagem de mina ácida, Tyson *et al.*, 2004; intestino humano, Kurokawa *et al.*, 2007).

Desde 2006, uma série de estudos relacionados à diversidade microbiana através de análise da região 16S do RNA ribossomal (Sogin *et al.*, 2006), viromas (Angly *et al.* 2006) e da genética ambiental microbiana (Rohwer *et al.*, 2006) foram conduzidos com a utilização de sequenciadores de larga escala. Abordagens baseadas nesses métodos de sequenciamento têm revolucionado a área de genética nos últimos anos, aumentando o número de sequências geradas e seu tamanho, enquanto que o preço vem caindo drasticamente. Essas plataformas de sequenciamento em larga escala são uma ferramenta poderosa e revolucionária para se obter rapidamente grandes quantidades de sequências de DNA, oferecendo um custo-benefício muito melhor do que o sequenciamento pelo método de Sanger. Atualmente estão disponíveis no mercado, as principais plataformas de sequenciamento de segunda geração: SOLiD (Applied Biosystems), Illumina/Solexa, Ion Torrent (Life Technologies) e o 454/Life Science. Dentre estes apenas a plataforma

454/Life Science é capaz de gerar sequências de até 1000 pb, através do sistema GS FLX + (<http://www.454.com/products/gs-flx-system>). Esse sistema do 454/Life Science facilita a montagem de sequências “contigs” (grupo de sequências sobrepostas de DNA que juntas representam uma região consenso de DNA), ou até “scaffolds” (são “contigs” sobrepostos separados por espaços de comprimento conhecido), aumentando as chances de se encontrar os genes e/ou proteínas de interesse. Como a metagenômica envolve o sequenciamento aleatório da informação genética de amostra ambiental, é inviável obter informações suficientes a partir de sequências pequenas (menores que 100bp) para caracterizar os segmentos de DNA em grupos funcionais ou em espécies (Krause *et al.*, 2008; Jones 2010). O sequenciamento de metagenomas, principalmente por pirosequenciamento (454/Life Science), permite estimar de uma só vez, em um dado momento, a presença de diferentes patógenos em amostras de água (vírus, bactérias, protozoários e fungos) e realizar a estimativa de abundância de genes de resistência a antibióticos, além de outros, de forma ampla e mais próxima da biodiversidade total da amostra. Entretanto, dados de sequenciamento de metagenomas por “Shotgun” (sequenciamento direto e aleatório de uma amostra ambiental, sem prévia amplificação de qualquer gene) estão entre os tipos de “informações biológicas” mais desafiadoras devido ao tamanho e complexidade da amostra (que dependerá do ambiente). Mesmo assim, atualmente essa abordagem representa o melhor método disponível para o estudo da vasta comunidade não caracterizada de micro-organismos (e seus genes funcionais) (Tanenbaum *et al.*, 2010).

Vários ambientes têm sido estudados por abordagens metagenômicas incluindo o solo, a cavidade oral, fezes, ambientes aquáticos, assim como esgotos hospitalares, também chamado de metagenoma hospitalar, que é um termo envolvendo o potencial genético de organismos encontrados em hospitais que preocupam as autoridades de saúde pública como a resistência a antibióticos e as infecções nosocomiais (Riesenfeld *et al.*, 2004, Coque *et al.*, 2002). Infelizmente não existe uma única base de dados que agrupe todas as sequências de projetos metagenômicos, sendo que a qualidade e descrição dos dados da coleta e projetos variam entre as bases de dados e entre os grupos de dados.

A análise metagenômica fornece uma metodologia para avaliar o potencial metabólico de certo ambiente, representando um retrato daquele dado momento no qual o DNA daquele ambiente pode fornecer uma visão ampla da comunidade microbiana em termos de taxonomia e potencial metabólico (Steele & Streit, 2005). Antes do sequenciamento em larga escala, os genes de resistência a antibióticos eram isolados de amostras ambientais através da amplificação por PCR ou pela clonagem de bactérias cultiváveis (Riesenfeld *et al.*, 2004). Porém, esses métodos não permitem analisar o potencial dos reservatórios de genes de resistência porque a grande maioria das bactérias não é cultivável e a detecção destes genes por PCR depende de iniciadores (“primers”) específicos baseados em genes de resistência já conhecidos. Sendo assim, essa técnica dificulta a descoberta de novos genes. O desenvolvimento de técnicas independentes de cultura facilitou o descobrimento de novos genes e o acesso à diversidade genética da maioria das bactérias (Schmieder & Edwards, 2012). Nesse contexto, a metagenômica é uma das abordagens mais modernas que permite superar as limitações dos métodos baseados em cultura ou amplificação (Hendelsman *et al.*, 1998) e permite descrever o potencial genético de uma comunidade e identificar tipos de micro-organismos presentes em uma comunidade, assim como a presença de genes ou variações genéticas responsáveis pela resistência a antibióticos. Através da metagenômica diversos genes novos de resistência a antibióticos já foram identificados, incluindo resistência a  $\beta$ -lactâmicos (Allen *et al.*, 2009; Donato *et al.*, 2010), tetraciclina (Diaz-Torres *et al.*, 2003), aminoglicosídeos (Riesenfeld *et al.*, 2004; Donato *et al.*, 2010) e bleomicina (Mori *et al.*, 2008).

### **1.7 Bases de dados metagenômicos e de resistência a antibióticos**

Apesar da resistência a antibióticos ser intensivamente estudada na área clínica, pouco se sabe sobre os reservatórios ambientais desses genes e sua contribuição para a resistência na clínica. A identificação de fontes de genes de resistência e seu monitoramento no ambiente irão contribuir para o avanço das medidas de combate à resistência a esses antibióticos em patógenos humanos (Allen *et al.*, 2009). Neste contexto, a metagenômica pode ajudar a acessar diretamente o conteúdo genômico da comunidade microbiana, possibilitando a identificação e anotação de genes ambientais, além de ajudar nos estudos de

metagenoma funcional, mesmo considerando que ainda é difícil detectar e prever com acurácia as funções a partir de sequências de tamanho médio (~400 – 500 pb). Muitas sequências ambientais já estão disponíveis em diferentes bases de dados como NCBI, CAMERA (“Community Cyberinfrastructure for Advanced Microbial Ecology Research & Analysis”, <http://camera.calit2.net/>), MG-RAST (“Metagenomic Rapid Annotation using Subsystem Technology”, <http://metagenomics.nmpdr.org>), e IMG/M (“Integrated Microbial Genomes”, <http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>). Em relação aos genes de resistência a antibióticos já existem na literatura bases de dados específicas como o ARDB (“Antibiotic Resistance Genes Database”) (Liu & Pop, 2008).

## 2 JUSTIFICATIVA

Os esgotos oriundos de hospitais são uma fonte importante de disseminação de patógenos e outros organismos para o meio ambiente aquático e sua posterior reinserção no ambiente social. Logo, esse tipo de efluente possui um papel importante na disseminação de genes de resistência no ambiente. Sem o devido tratamento, esses esgotos se tornam graves problemas de saúde pública por serem grandes responsáveis pelo lançamento de micro-organismos patogênicos portadores de genes de resistência aos antimicrobianos no ambiente (Gusatti *et al.*, 2009). Por esta razão, é necessário conhecer as características desse efluente (presença de genes e micro-organismos nocivos ao homem) na tentativa de estimar o risco ambiental e para a saúde pública de um lançamento acidental ou irregular de esgoto hospitalar em ambientes aquáticos como um rio, lagoa ou o mar.

Este projeto tem como foco explorar e estudar a diversidade molecular das enzimas relacionadas à resistência de  $\beta$ -lactâmicos ( $\beta$ -lactamases) de dois esgotos hospitalares localizados na cidade do Rio de Janeiro, assim como de estudos de metagenomas disponibilizados em bases públicas. Para isso, foram selecionados dois hospitais públicos por serem frequentados por um grande número de pessoas, principalmente a porção mais carente que vive em condições sanitárias precárias, sendo uma importante fonte de disseminação dos genes de interesse. Para estudar a diversidade dessas enzimas presentes nos esgotos hospitalares foi utilizada a abordagem de metagenômica através do sequenciamento aleatório direto (“Shotgun”). Poucos estudos foram realizados até o momento empregando abordagens metagenômicas para estimar a presença patogêna e resistência aos antibióticos em diferentes ambientes (Allen *et al.*, 2009; Kazimierczak *et al.*, 2009; Donato *et al.*, 2010; Torres-Cortés *et al.*, 2011). Entretanto, esses poucos estudos têm mostrado que esse tipo de abordagem apresenta um grande potencial para a descoberta de novos genes, levando a um aumento constante do número de projetos de metagenoma que utiliza pirosequenciamento (Gill *et al.*, 2006). A grande vantagem dessa abordagem é o grande volume de sequências geradas a um baixo custo em comparação com o sequenciamento via tecnologia Sanger, além de não apresentar o viés da clonagem uma vez que o DNA ambiental é

sequenciado diretamente. Apesar do grande potencial, esta metodologia ainda não foi empregada para avaliar especificamente a presença ou o impacto de GRAs no meio ambiente, mais precisamente a presença de genes das  $\beta$ -lactamases em amostras de esgotos hospitalares no Estado do Rio de Janeiro. O grande número de genes de resistência já descritos torna as abordagens moleculares, como PCR, pouco eficientes. Em contraste, a metagenômica direta fornece uma alternativa exploratória de culturas independentes de cultivo de forma a caracterizar a genética de resistência dentro das comunidades microbianas (Kristiansson *et al.*, 2011), permitindo o estudo de genes de resistência. As pequenas enzimas e moléculas já descobertas indicam o grande potencial da metagenômica na mineração de genes de interesse biotecnológico e no conhecimento da diversidade genética e de comunidades microbianas.

## **3 OBJETIVOS**

### **3.1 Geral**

Identificar e classificar homólogos de  $\beta$ -lactamases em estudos de metagenoma disponíveis em bases públicas e nas amostras de esgotos hospitalares do Rio de Janeiro (EHRJ) pirosequenciadas, com foco na classificação dessas enzimas e análise de sua diversidade filogenética.

### **3.2 Específicos**

- Minerar  $\beta$ -lactamases em (a) metagenomas ambientais dispostos em bases públicas e (b) amostras de EHRJ utilizando perfis HMM.
- Caracterizar o perfil taxonômico da comunidade microbiana das amostras de EHRJ geradas pelo pirosequenciamento;
- Caracterizar o potencial metabólico da comunidade de Bacteria presente nas amostras de EHRJ geradas pelo pirosequenciamento;
- Analisar a diversidade genética das  $\beta$ -lactamases encontradas nos EHRJ e nos projetos de metagenoma públicos;

## **4 METODOLOGIA**

A metodologia desta tese está dividida em duas seções. A primeira descreve os métodos empregados na parte experimental que consiste em coletas, análises físico-químicas e biologia molecular, e a segunda consiste nas análises realizadas “in silico”, subdividida em duas partes: (i) busca por prováveis homólogos de  $\beta$ -lactamases em bases de dados metagenômicos e na base de esgotos hospitalares, e análise filogenética dessas enzimas identificadas, e (ii) caracterização taxonômica e funcional dos dados gerados pelo pirosequenciamento dos esgotos hospitalares do Rio de Janeiro (EHRJ).

### **PARTE 1: Parte experimental**

#### **4.1 Escolha dos locais de coleta**

Optou-se por realizar a coleta de esgotos de diferentes hospitais públicos, envolvendo diferentes comunidades para tentar ampliar o acesso à diversidade de  $\beta$ -lactamases. Sendo assim, cartas solicitando a permissão para a inclusão dos hospitais no projeto da tese (implicando na realização de coletas dos seus esgotos) foram redigidas e entregues em hospitais públicos localizados nas regiões da Zona Sul e Zona Norte da cidade do Rio de Janeiro – RJ, Brasil. Os hospitais selecionados foram os dois primeiros que responderam e concederam as autorizações para as coletas das amostras dos esgotos.

#### **4.2 Coleta da amostra (efluente hospitalar)**

Foram coletados 20 litros de amostras de esgoto de cada um dos dois hospitais selecionados. A coleta no hospital da Zona Sul foi realizada no dia 27 de abril de 2011, sendo que o esgoto bruto foi retirado de um poço que recebia esgoto das áreas de cardiologia, ortopedia, clínica médica e laboratório de análises clínicas. Era a região de mais fácil acesso a qual se conseguiu a permissão. Esse hospital tem como característica ser um hospital de grande porte, cobrindo a área do Município do Riode Janeiro, com um total de 400



leitos e com um total de 13.273 internações em 2011. É mais conhecido como um hospital de emergência.

No hospital da Zona Norte a coleta foi realizada dia 26 de maio de 2011, sendo que lá o sistema de esgoto bruto não é tão compartimentalizado quanto o primeiro hospital, então as amostras coletadas contemplavam todo o hospital, uma vez que o poço recebia esgotos de todas as áreas. Esse hospital é caracterizado por ser universitário, possuindo diversos tipos de tratamentos, inclusive de câncer, possuindo um total de 525 leitos e um total aproximado de 18.000 internações em 2011. Nos dois hospitais as caixas de esgoto bruto foram indicadas e sua abertura acompanhada por funcionários do setor responsável pela manutenção de cada hospital trabalhado. As coletas foram realizadas com utilização de EPIs (equipamentos de proteção individual) apropriados para esta tarefa. Ambas as coletas foram realizadas em período diurno. Após a coleta, as mesmas foram resfriadas a aproximadamente 4 °C utilizando-se gelo e levadas imediatamente ao laboratório. Um litro de cada amostra foi deixado no DSSA (Departamento de Saneamento e Saúde Ambiental), localizado na ENSP (Escola Nacional em Saúde Pública)/ FIOCRUZ, para a realização das análises físico-químicas e colimetria, como descrito no item 4.3. O restante foi levado ao laboratório LBCS onde foi submetido à filtração.

As amostras foram filtradas com auxílio de bombas a vácuo, primeiro utilizando um pré-filtro médio (0,5mm  $\varnothing$  47  $\pm$  0,5 mm – Schleicher & Schuell) para reter as partículas sólidas. Posteriormente foram filtradas em membranas de diferentes porosidades como 1,2  $\mu$ m, 0,8  $\mu$ m e 0,22  $\mu$ m (1,2  $\mu$ m e 0,8  $\mu$ m  $\varnothing$  47  $\pm$  0,5 mm - Schleicher & Schuell; e 0,22  $\mu$ m  $\varnothing$  47 mm – Millipore TM). A utilização de várias membranas de diferentes porosidades teve como objetivo não permitir que a membrana entupisse e houvesse uma perda maior das células de interesse, já que os esgotos hospitalares costumam conter muitas partículas sólidas, tornando-se mais denso. A primeira membrana (1,2  $\mu$ m) foi utilizada para reter o material sólido, a segunda (0,8  $\mu$ m) para reter a maior parte dos microeucariotos e a terceira (0,22  $\mu$ m), de menor porosidade, foi utilizada para reter a maior parte das células procarióticas. Após a filtração as membranas foram cortadas ao meio com o auxílio de lâminas de bisturi e pinças estéreis, e estocadas em tubos Falcon a -80°C, até sua utilização. Mesmo após a filtração do efluente pelo pré-filtro e pelas membranas de 1,2 e

0,8 µm, foi observado entupimento da membrana de 0,22 µm após a passagem de 100 mL do líquido. Logo, para ter uma padronização, cada membrana de 0,22 µm continha aproximadamente 100 mL da amostra de efluente.

### **4.3 Análise físico-química e colimetria**

As análises físico-químicas e de colimetria foram realizadas pelo grupo do DSSA, ENSP/FIOCRUZ, conforme descrito por Chagas *et al* (2011). Os parâmetros físico-químicos como temperatura da água, pH e demanda bioquímica de oxigênio após 5 dias (DBO<sub>5</sub>) foram medidas através de uma sonda multiparamétrica (Hanna Instruments modelo HI 9828) inserida na água (efluente). As concentrações de fósforo e nitrogênio total foram obtidas através de método colorimétrico, usando o kit Hach Nitruver CN66, e a análise da demanda química de oxigênio (DQO) foi avaliada pelo método padrão de dicromato de potássio como agente oxidante. A análise de coliformes totais e fecais foi realizada através do método de fermentação em tubos múltiplos, padronizada pelo “Standard Methods for Examination of Water and Wastewater” (APHA, 1998). As células bacterianas foram quantificadas utilizando o kit Colilert P/A Quanti tray 2000 (Laboratories, Westbrook, ME, EUA), de acordo com as instruções do protocolo da empresa. As amostras foram coletadas em frascos plásticos, esterilizados, previamente limpos em água destilada ou deionizada, sendo encaminhadas ao laboratório em menos de uma hora após a coleta. As colônias bacterianas foram contadas após 18h de incubação a 37° C. As células que apresentaram uma coloração amarela representam os coliformes totais e as células fluorescentes (após a emissão de luz ultravioleta) representam as *E. coli*. Os números de coliformes foram obtidos através da Tabela do Número Mais Provável (NMP) e representam os valores encontrados em 100 mL de amostra (NMP/100 mL).

### **4.4 Extração de DNA metagenômico**

Os procedimentos para a extração de DNA foram realizados no laboratório de Biologia Molecular de Parasitos e Vetores do IOC-FIOCRUZ.

Para a extração do DNA das membranas foi utilizado o kit de extração “Metagenomic DNA isolation Kit for Water” (EPICENTRE Biotechnologies, Madison, WI, EUA) de acordo com as recomendações do fabricante. A quantificação e medida de parâmetros de qualidade do DNA (como a razão 260/280 nm) foi realizada através da leitura no espectrofotômetro Nanodrop (ND 1000 Spectrophotometer V3.7, NanoDrop Technologies) e também através de eletroforese em gel de agarose 1% (p/v), em tampão TAE 1X (Tris-acetato-EDTA) adicionado de brometo de etídio (0,5 µg/mL), utilizando diferentes concentrações (5ng, 10 ng, 20ng e 50ng) do marcador de peso molecular de 40 Kb (presente no kit de extração). Após a corrida de aproximadamente 1h, o gel foi visualizado sob luz UV e registrado através de um sistema de fotodocumentação de géis (MiniBis Pro, Bioimages System) e a quantificação foi realizada pela comparação da intensidade das bandas no gel.

Uma alíquota contendo aproximadamente 2µg do DNA extraído foi utilizada para a reação de pirosequenciamento (descrito no item 4.5.5) e uma alíquota menor (~5 ng) para a amplificação do 16S ribossomal (rrs) por PCR, conforme descrito adiante. Para a obtenção dos 2µg de DNA metagenômico, quantidade necessária para o pirosequenciamento, extraiu-se o DNA a partir de nove membranas (de 0.22µm) da amostra de cada hospital (~900 mL). Como foram realizadas várias extrações e o volume final ficou alto (50 µL após cada extração), o DNA de cada amostra foi misturado e concentrado no *Speedvack Concentrator Plus* (Eppendorf) (3.000 rpm por aproximadamente 30 min) de modo a obter a concentração necessária para os procedimentos de sequenciamento.

Vale ressaltar que existem duas réplicas para cada hospital. Desta forma, as extrações das réplicas foram realizadas em separado, gerando quatro amostras de DNA no final. As réplicas 01 e 02 do hospital da Zona Sul (ZS01 e ZS02) e as réplicas 01 e 02 do hospital Zona Norte (ZN01 e ZN02).

#### **4.5 Caracterização do DNA metagenômico**

As amostras foram submetidas a duas formas de sequenciamento: primeiro usando o método de Sanger com o objetivo de uma análise preliminar da diversidade da comunidade bacteriana usando um gene ribossomal, e depois usando o pirosequenciamento de biblioteca de DNA metagenômico

total (denominada “Shotgun”) visando explorar a diversidade taxonômica, mas também permitir a busca e exploração da diversidade dos genes de interesse ( $\beta$ -lactamases). Para o sequenciamento pelo método de Sanger foram utilizados fragmentos de 16S ribossomal (rrs) gerados após a amplificação pela técnica de PCR, a partir de uma alíquota do DNA metagenômico extraído das amostras.

#### **4.5.1 Amplificação de uma região do gene ribossomal 16S**

Uma alíquota do DNA extraído das réplicas 01 e 02 de cada amostra de efluente hospitalar (conforme descrito no item 4.4) foram unidas permanecendo apenas duas amostras, ZN e ZS. A partir do DNA extraído foram gerados fragmentos do gene 16S ribossomal (rrs) de Bacteria utilizando-se iniciadores universais BAC27F (5'-AGAGTTTGATCMTGGCTCAG-3') e BAC518R (5'-ATTACCGCGGCTGCTGG-3'). A amplificação do gene 16S ribossomal (rrs) de Bacteria foi feita de acordo com o seguinte protocolo: mix contendo 2,5  $\mu$ L de tampão para PCR 10x; 0,2 mM de dNTP; 3,0 mM de MgCl e 1U de Taq DNA polimerase recombinante (Fermentas); 1  $\mu$ L da amostra de DNA (contendo aproximadamente 4ng de DNA); 5 pmol de cada primer; água Milli-Q esterilizada para um volume final de 25  $\mu$ L, nas seguintes condições: 95° C por 5 minutos; 30 ciclos; 92°C por 1 minuto, 55°C por 1 minuto e 72°C por 1 minuto; 72°C por 10 minutos.

#### **4.5.2 Clonagem das bibliotecas de 16S ribossomal e sequenciamento pelo método de Sanger**

Os “amplicons” obtidos (com aproximadamente 500 pb) foram retirados do gel (com a ajuda de uma lâmina de bisturi assepsiada com álcool 70%), purificados através do QIAquick Gel Extraction Kit (Qiagen, Hilden, Alemanha) de acordo com as instruções do fabricante e clonados utilizando o kit pGEM<sup>®</sup> T Easy Vector System (Promega, Madison, WI, EUA). Os clones obtidos foram transformados em células quimiocompetentes de *Escherichia coli* DH5 $\alpha$  através de choque térmico (30 min no gelo, seguido por 45 seg a 42°C e 2 min no gelo), plaqueadas em meio LB-agar contendo ampicilina (100 $\mu$ g mL<sup>-1</sup>), X-Gal (1%) e IPTG (10mM) e incubadas durante a noite a 37°C ( $\pm$  18h). As colônias contendo o inserto foram retiradas com a ajuda de palitos estéreis e adicionados em placas tipo “deep well” (poço profundo com fundo reto), de 96

poços, contendo meio líquido LB e ampicilina ( $100 \mu\text{g mL}^{-1}$ ). Após esta etapa as placas foram incubadas por 18h a  $37^{\circ}\text{C}$ . Após o crescimento, o DNA foi extraído por um método de lise alcalina, conforme descrito por Sambrook & Russell (2001). Após esta etapa o DNA foi sequenciado com auxílio do equipamento ABI-3730 (Applied Biosystems) da plataforma de sequenciamento PDTIS-FIOCRUZ utilizando o kit “BigDye Terminator v3.1” (Applied Biosystems).

#### **4.5.3 Pré-processamento das sequências de 16S ribossomal geradas pelo método de Sanger**

Após o sequenciamento dos clones da biblioteca de 16S ribossomal, os eletroferogramas foram processados pelo programa PHRED (Ewing *et al.*, 1998) para realizar a definição de base (“base calling”), a remoção de bases de baixa qualidade ( $> 20$ , menos de um erro em 100 nucleotídeos) e as regiões das sequências correspondentes ao vetor, assim como as sequências com tamanho inferior a 200 bp, através sistema de análise Stingray (stingray.biowebdb.org). Somente as sequências com qualidade alta ( $Q > 20$ ) foram utilizadas nas análises posteriores. A presença de sequências quiméricas foi avaliada com o auxílio do programa Decipher (parâmetro padrão, versão outubro de 2012) (<http://decipher.cee.wisc.edu/>) e estas foram removidas das análises posteriores.

#### **4.5.4 Estimativa de riqueza e diversidade da comunidade bacteriana utilizando parte do gene ribossomal 16S**

Para as sequências de 16S ribossomal (rrs) geradas pelo método de Sanger, as unidades taxonômicas operacionais (UTOs) foram agrupadas utilizando o programa MOTHUR 1.22.2 (“Distance-based Operational Taxonomy Unity and Richness Determination”) usando um ponto de corte de distância evolutiva de 0,03 (Schloss *et al.*, 2009) para espécies. Para isso as sequências obtidas após o pré-processamento foram primeiro alinhadas com o programa ClustalW 1.82 (utilizando os parâmetros padrão) (Thompson *et al.*, 1994) e matrizes de distâncias foram construídas com o programa DNADIST do pacote PHYLIP 3.6 (Felsenstein, 1993), utilizando os parâmetros padrão e o modelo de substituição Jukes-Cantor (Jukes & Cantor, 1969). As matrizes geradas foram utilizadas como arquivo de entrada para o programa MOTHUR para

calcular o agrupamento das seqüências em UTOs, bem como o índice de riqueza das espécies através dos métodos não paramétricos Chao1 (Chao, 1987) e, os índices de diversidade através dos métodos de Shannon (Shannon & Weaver, 1949) e ACE (Chao & Lee, 1992), além da curva de rarefação. O estimador de riqueza Chao 1 calcula o número mínimo de OTUs presentes em uma amostra. Já o índice de Shannon calcula o grau médio da incerteza da predição de quais espécies de um indivíduo escolhido ao acaso de uma coleção de S espécies e N indivíduos pertence. O valor aumenta quando o número de espécies aumenta e quando a distribuição dos indivíduos entre as espécies se iguala.

#### **4.5.5 Classificação taxonômica**

Após a obtenção das seqüências de 16S ribossomal, estas foram caracterizadas taxonomicamente através do classificador (“classifier”) presente na base de dados RDP (“Ribossomal Database Project”, <http://rdp.cme.msu.edu/classifier.jsp>).

#### **4.5.6 Pirosequenciamento das amostras**

Para o pirosequenciamento foram utilizadas amostras de esgotos hospitalares (descrito no item 4.2) contendo aproximadamente 2 µg de DNA. O pirosequenciamento foi realizado no Laboratório Nacional de Computação Científica (LNCC), localizado na cidade de Petrópolis (RJ) através da plataforma 454/Life Science com uso do sistema GS FLX+. Esse sistema gera seqüências de até 1000 pb de comprimento, com um perfil de 85% das seqüência com tamanho acima de 500pb e 45% do total acima de 700pb ([454.com/products/gs-flx-system/index.asp](http://454.com/products/gs-flx-system/index.asp)). Foram realizadas duas corridas, sendo que em cada corrida a placa foi dividida em duas partes contendo uma réplica biológica da coleta em cada uma. Na primeira corrida foi utilizado DNA dos esgotos hospitalares da Zona Sul (ZS) e na segunda corrida o DNA extraído dos esgotos hospitalares da Zona Norte (ZN). Na primeira corrida foram utilizadas as réplicas biológicas 01 e 02 do hospital ZS, enquanto que na segunda corrida foram utilizadas as réplicas 01 e 02 do hospital ZN.

## **PARTE 2- Parte “In silico”**

### **4.5.7 Pré-processamento das sequências geradas pelo aparelho 454/Life Science**

As sequências foram extraídas dos arquivos “Standard flowgram format” (SFF) utilizando o conversor “sff extract” (versão 0.28) que é oferecido, de forma gratuita, pela própria plataforma 454/Life Science. A partir daí foram gerados dois arquivos, “fasta” e “qual”, carregadas no banco de dados Stingray (<http://stingray.biowebdb.org>) que se baseia no esquema “Genomics Unified Schema” (GUS). Em seguida, as sequências foram filtradas em duas etapas: a primeira etapa foi a remoção de artefatos gerados pela tecnologia do sequenciamento, através do programa CD-HIT-454 (versão 4.5.4, parâmetro “default” com a opção “- c 0.98” (98% de identidade) como ponto de corte (Beifang *et al.*, 2010); a segunda etapa foi a filtragem por tamanho (>100pb) e qualidade (= ou > 20) com o programa Lucy (versão 1.2) com parâmetros padrão (Chou & Holmes, 2001).

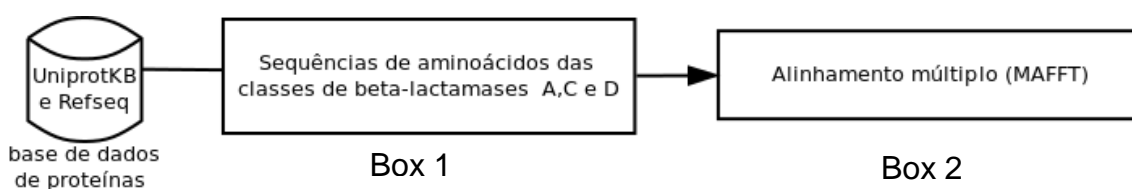
### **4.6 Construção de pHMMs para a mineração de homólogos de $\beta$ -lactamases**

#### **4.6.1 Seleção de sequências “sementes” para a construção de alinhamentos**

Para a construção dos perfis HMM (“Hidden Markov Model”) para cada classe de  $\beta$ -lactamases foram obtidas sequências curadas da base de dados UniprotKB/Swiss-Prot/Trembl (<http://www.ebi.ac.uk/uniprot>) e Refseq. As sequências obtidas do Refseq foram de enzimas que tiveram sua função confirmada experimentalmente. Essas sequências foram utilizadas como referência, chamadas de proteínas “sementes”. No total foram utilizadas 36 sequências de aminoácidos de  $\beta$ -lactamases da classe A, 30 sequências da classe C e 30 da classe D. Os números de acesso dessas sequências se encontram no Apêndice.

#### **4.6.2 Alinhamento múltiplo de sequências**

Os alinhamentos das sequências “sementes” assim como os outros alinhamentos utilizados nesta tese foram construídos usando o MAFFT versão 6.717b com o parâmetro “--auto” (Kato & Kuma, 2002). A visualização dos alinhamentos foi feita com o programa Jalview versão 2 (Waterhouse *et al.*, 2009). Nos alinhamentos múltiplos das sequências “sementes” foram retiradas as regiões de peptídeo sinal que são muito polimórficas. As sequências as quais se desconhecia a posição (início e fim) do peptídeo sinal foram comparadas no alinhamento múltiplo com aquelas que se sabia a posição correta, e um corte foi realizado em todas as sequências alinhadas, a partir daquelas em que a posição era conhecida. A Figura 4.2 ilustra as etapas 4.6.1 e 4.6.2.



**Figura 4.1** Box 1: Base de dados de proteínas de onde foram retiradas as sequências de aminoácidos das  $\beta$ -lactamases das classes A, C e D para a construção dos perfis HMMs de cada classe. Box 2: Alinhamento múltiplo das sequências de proteína das classes A, C e D de  $\beta$ -lactamases com o programa MAFFT para posterior construção dos pHMMs referentes a cada classe.

#### 4.6.3 Perfis de Cadeia Oculta de Markov (pHMM)

O programa HMMER (“Profile Hidden Markov Models for biological sequence analysis”), versão 3.0 (março de 2010) foi utilizado para a construção de perfis HMM e realizar as buscas de sequências homólogas às proteínas que originaram os perfis em bases de dados. Ele implementa métodos utilizando modelos probabilísticos chamados de “Perfis ocultos de Markov” (“Hidden Markov Models”), também conhecidos como perfis HMM (Eddy, 1998; Finn *et al.*, 2011). Os programas utilizados do pacote HMMER foram dois:

“**hmmbuild**” – constrói modelos probabilísticos que representam uma família de sequências de proteína.

“**hmmsearch**” - busca sequências de proteína em bases de dados a partir de um pHMM para cada classe. O programa lê o pHMM e faz a busca em uma determinada base de dados por sequências similares, dentro de um ponto



de corte (“E-value”) a ser determinado pelo usuário. Os resultados são apresentados em uma lista de forma ordenada e decrescente das sequências que apresentaram os melhores “E-values” (menores valores). Nesta tese foram utilizados os parâmetros padrão para o “hmmbuild” e o “hmmsearch”, o ponto de corte utilizado foi “E-value”  $10^{-5}$  para buscar  $\beta$ -lactamases também mais divergentes.

#### **4.6.4 Construção dos perfis de Cadeia Oculta de Markov (pHMM)**

As sequências retiradas da base UniprotKB (maioria curada, ou seja, manualmente anotada e revisada) e Refseq (item 4.6.1) de cada classe de serina  $\beta$ -lactamases (A, C e D) foram alinhadas separadamente para a criação de alinhamentos “semente” e a presença do sítio catalítico e motivos específicos foram verificados manualmente, utilizando o programa Jalview versão 2 (Waterhouse *et al.*, 2009), conforme descrito no item 4.6.2. Os alinhamentos obtidos com o MAFFT 6.717b (Kato & Kuma, 2002) foram formatados para “Stockholm”, que é o formato aceito pelo HMMER 3, através de um “script” na linguagem “perl”, e utilizados para a construção de pHMM através do programa “hmmbuild”. Esses perfis foram utilizados para a busca de homólogos de  $\beta$ -lactamases nas bases públicas CAMERA e IMG/M (item 4.7.1), e também no grupo de sequências referente às amostras de esgotos hospitalares do Rio obtidas de acordo com o item 4.2. Além desses perfis iniciais (conforme descrito no item 4.6) foram construídos também pHMMs dos três motivos principais de cada classe. Esses motivos são: SxxK, SDN e KTG (classe A), SxSK, YSN e KTG (classe C) e SxxK, YGN e KTG (classe D). Para isso, a partir do alinhamento múltiplo “semente” referente a cada classe de  $\beta$ -lactamases, a região de cada motivo foi cortada, incluindo cinco aminoácidos posteriores e anteriores a esta região, e foi realizada a construção de pHMMs desses motivos específicos utilizando o programa “hmmbuild” do pacote HMMER.

Após a construção dos perfis, estes foram visualizados graficamente através da ferramenta LogoMat-M, disponível no portal do Instituto Sanger (<http://www.sanger.ac.uk/resources/software/logomat-m/>). O logo é representado por pilhas de letras de diferentes tamanhos. Para cada estado de

emissão do perfil, uma pilha de letras é formada. A altura das pilhas é determinada pelo desvio da posição da frequência de emissão da posição das letras das frequências em segundo plano. A altura das letras representa a probabilidade de atingir um estado e do número esperado de letras que o estado emite durante sua passagem no modelo de Markov (contribuição esperada do estado). O perfil da cadeia oculta de Markov especifica a distribuição da emissão de uma letra em uma posição específica, e também a posição específica das probabilidades de inserção e deleção para descrever a sequência de uma família. Ou seja, o modelo emite uma letra do alfabeto cada vez que um estado é visitado.

#### **4.7 Busca por homólogos de serina $\beta$ -lactamase em bases de dados metagenômicos**

##### **4.7.1 Bases de estudo**

Duas bases de dados públicos foram utilizadas nesse estudo. A primeira foi o CAMERA versão 2.0 (<http://câmera.calit2.net/>) que é um repositório de dados metagenômicos iniciado com o projeto GOS (*Global Ocean Sampling Expedition*) realizado em diferentes pontos do Mar de Sargasso, localizado no Triângulo das Bermudas. Esse projeto iniciou em 2004 e as sequências foram geradas a partir do método de sequenciamento de Sanger. A partir de 2005 os dados depositados foram obtidos pelo método de pirosequenciamento através do 454/Roche. Atualmente existem 84 projetos disponíveis no CAMERA, de diferentes tamanhos, indo de 1.2 Mb (menor) até o projeto GOS que tem 20 Gb (maior). Após o banco ter sido traduzido nos seis quadros de leitura, utilizando o programa TRANSEQ do pacote EMBOSS (v6.1.0), este apresentou 75 Gb e 1.052.288.094 de sequências. A segunda base de dados foi o IMG/M versão 3.0 (<http://img.jgi.doe.gov/cgi-bin/m/main.cgi>) que contém 364 projetos, de tamanhos que variam entre 20 Kb (menor) e 457 Mb (maior) contendo sequências em formato de aminoácidos, totalizando 20Gb e 119.059.610 sequências de aminoácidos. Estes projetos envolvem diferentes ambientes naturais como terrestre aquático e ar, os associados a diferentes hospedeiros (mamíferos, artrópodes, humanos, plantas, moluscos, aves) e os “engenheirados” como esgoto, material de biorremediação, de

biotransformação, de enriquecimento e de resíduos sólidos. Os dados foram obtidos em junho de 2012.

#### 4.7.2 Busca por homólogos de $\beta$ -lactamases

A partir dos pHMMs construídos conforme descrito no item 4.7 foi realizada uma busca das  $\beta$ -lactamases nos bancos ambientais do CAMERA e do IMG/M através do programa “hmmsearch”. Estes mesmos pHMMs foram utilizados para minerar as  $\beta$ -lactamases nos dados gerados do metagenoma dos esgotos hospitalares, conforme descrito posteriormente, no item 4.10. O esquema a seguir da Figura 4.3 ilustra tais etapas.

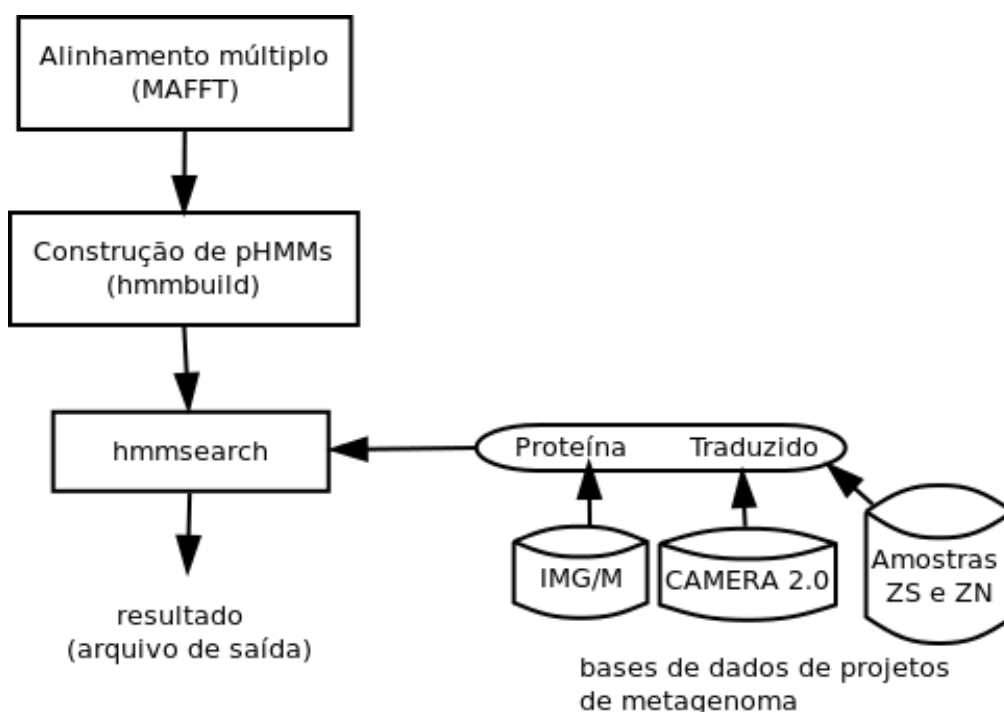


Figura 4.2. Esquema da metodologia relacionada aos itens 4.6.2, 4.6.3 e 4.7.2.

#### 4.8 Anotação funcional através da presença de domínios conservados por similaridade

Para confirmar a função das prováveis  $\beta$ -lactamases recuperadas pelo programa “hmmsearch”, as sequências de entrada de proteínas representando os correspondentes “hits” obtidos a partir dos itens 4.7.2 e 4.10 foram utilizadas

como entrada na busca por similaridade com bases de matrizes de posição específica de pontuação ou “Position-Specific Scoring Matrices” (PSSM), de domínios conservados, o CDD (“Conserved Domains Database”, versão 2.25, 2.4GB) que engloba as bases Pfam (verão 24.0 e tamanho de 566.6Mb), TigrFam 9.0 (312.7Mb), Smart v5.1 (27.4Mb), Prk v5.0 (901.6Mb) e COG (“Cluster of Orthologous Group”) (versão 1.0 e tamanho de 321.4 Mb). O programa utilizado para essa análise foi o RPS-BLAST, versão 2.2.21 (Marchler-Bauer *et al.*, 2002) e o ponto de corte adotado foi “E-value” de  $10^{-5}$ , que reflete o quanto esse alinhamento é esperado para ocorrer ao acaso. As bases de dados de domínios conservados foram obtidas do NCBI (<http://www.ncbi.nlm.nih.gov/>). Essa metodologia também foi utilizada para selecionar as sequências analisadas na reconstrução filogenética, uma vez que as sequências que se apresentaram como melhor “hit” de regiões conservadas de  $\beta$ -lactamase foram utilizadas para a análise filogenética, conforme descrito mais adiante no item 4.12.

#### **4.9 Construção de pHMM de regiões específicas de $\beta$ -lactamases**

A partir das sequências resultantes das buscas por prováveis homólogos com os pHMMs iniciais (item 4.7.2 e 4.10) e que apresentaram anotação funcional de  $\beta$ -lactamases após a busca com RPS-BLAST (item 4.8), estas foram utilizadas como base para as buscas posteriores com os perfis construídos para regiões específicas (motivos) de cada classe de  $\beta$ -lactamase. A primeira busca foi realizada com o perfil referente ao primeiro motivo (o sítio catalítico S-x-x-K para as classes A e D e S-x-S-K para a classe C) contra os resultados obtidos a partir da primeira busca no CAMERA e IMG/M. Em seguida, os arquivos “fasta” das sequências resultantes foram recuperados a partir do programa “fastacmd” do pacote BLAST (versão 2.2.21). As sequências resultantes foram utilizadas como base para a busca utilizando o segundo perfil, referente ao segundo motivo, o Ser-Asp-Asn [S-D-N] (para classe A), Tyr-Ser(Ala)-Asn [Y-S(A)-N] (para a classe C) e Tyr-Gly(Phe)-Asn [Y-G(F)-N] (para a classe D), e o último perfil, da região Lys-Thr-Gly [K-T-G], com suas substituições existentes, apresentadas na Tabela 1.1. No final, os arquivos “fasta” resultantes continham todas as sequências prováveis homólogas de  $\beta$ -lactamases contendo os principais perfis de cada classe. Os alinhamentos

múltiplos gerados também foram utilizados para a análise filogenética descritas no item 4.12.

#### **4.10 Busca por homólogos de serina $\beta$ -lactamases em bases de esgotos hospitalares**

A busca por homólogos de  $\beta$ -lactamases foi feita a partir de buscas pelo programa “hmmsearch”, utilizando os perfis específicos de cada classe, construídos conforme descrito no item 4.9. A busca foi realizada com os pHMMs (classes A, C e D) contra as sequências obtidas das amostras de esgotos hospitalares.

#### **4.11 Cálculo da abundância relativa dos homólogos de serina $\beta$ -lactamases**

A partir dos resultados obtidos nos itens 4.7.2 e 4.10 que apresentaram anotação de serina  $\beta$ -lactamase foram utilizadas para o cálculo da abundância relativa (%) de cada classe de  $\beta$ -lactamase em relação ao número total de sequências nucleotídicas obtidas do sequenciamento de cada amostra hospitalar (ZS e ZN) e de número de sequências nucleotídicas totais presentes em cada projeto de metagenoma da base CAMERA. As sequências da base do IMG/M não foram utilizadas para esse cálculo, pois estas estão dispostas em forma de proteínas oriundas de prováveis ORFs (“Open Read Frames”) obtidas a partir das sequências montadas em “contigs” de cada projeto, não sendo possível o cálculo de abundância da mesma forma. Por outro lado, a comparação com os dados das outras bases (CAMERA, ZN e ZS) é viável.

Além do cálculo da abundância relativa de cada classe de  $\beta$ -lactamase (A, C e D) suas subclasses foram classificadas a partir de busca por similaridade de regiões conservadas utilizando perfis PSSM construídos a partir de tipos de resistência (RPS-BLAST) disponíveis no ARDB (“Antibiotic Resistance Database”), sendo o valor de corte utilizado “E-value”  $10^{-5}$ .

#### **4.12 Análise da diversidade filogenética dos homólogos de serina $\beta$ -lactamases em bases ambientais e obtidas por pirosequenciamento**

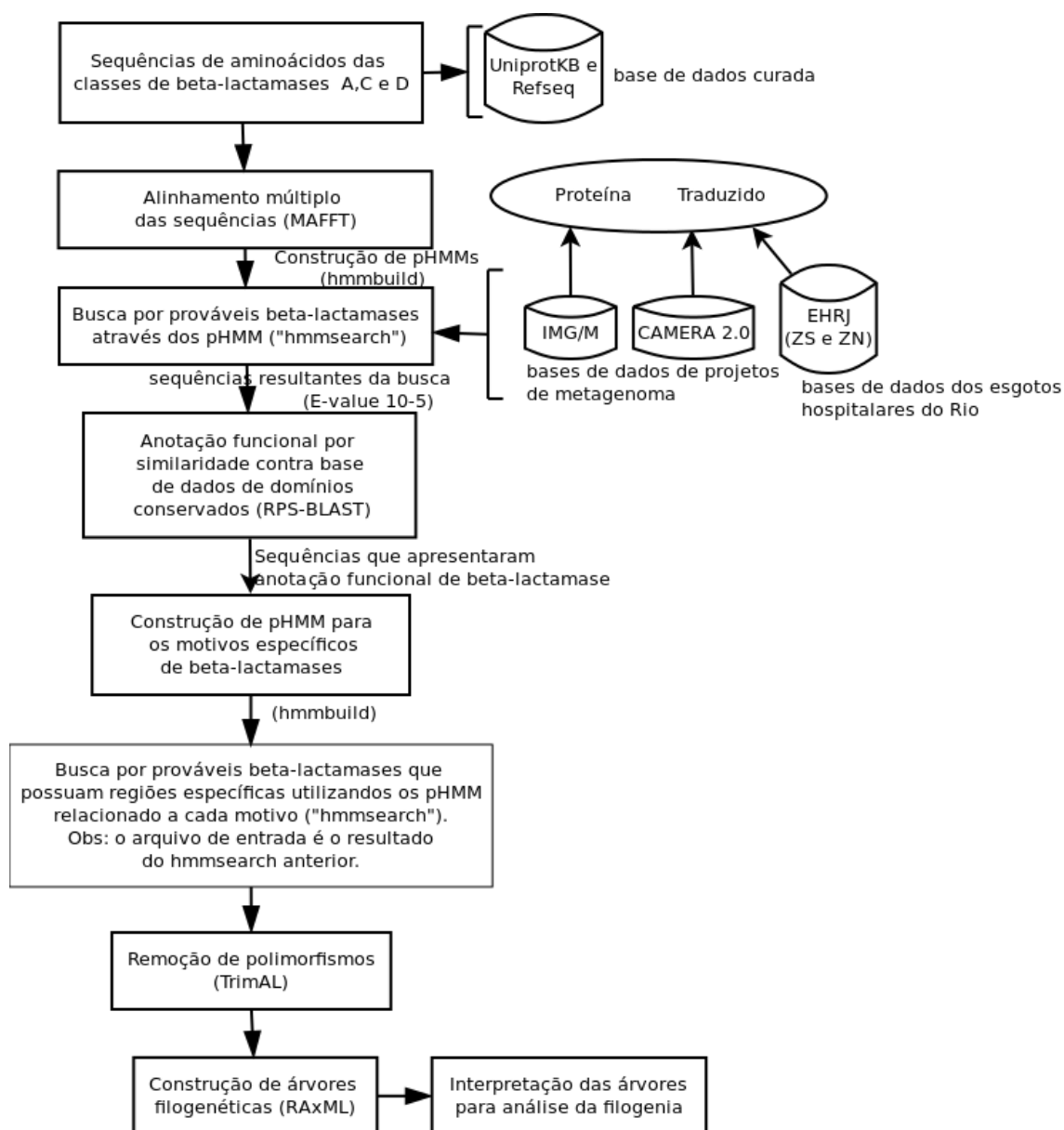
Após a obtenção das sequências pelos pHMMs de  $\beta$ -lactamases (conforme descrito no item 4.6.4) as sequências resultantes foram utilizadas para a análise filogenética. Para essa análise foram utilizados em conjunto os prováveis homólogos de  $\beta$ -lactamases obtidos nos itens 4.7 e 4.10, quatro sequências de  $\beta$ -lactamases de uma classe distinta (classe A ou D), usando como “outgroup” (grupo externo) e sequências representativas (“vizinhos”). Essas sequências representativas foram obtidas por análise de similaridade das sequências obtidas nos itens 4.7 e 4.10 contra a base de dados Refseq de proteína do NCBI (2780 Mb, março de 2012) utilizando o programa BLASTP (versão 2.2.21). Essas sequências foram reunidas e utilizadas na geração do alinhamento múltiplo utilizando o programa MAFFT versão 6.717b (Kato & Kuma, 2002), com a opção “--maxiterate 1000 --localpair in > out” que possui uma acurácia maior. Para as  $\beta$ -lactamases de classe A e C foram utilizados como grupo externo sequências da classe D e, para a classe D, foram utilizadas sequências da classe C. Como as sequências de aminoácidos das  $\beta$ -lactamases podem divergir bastante mesmo dentro das classes optou-se por identificar e retirar as posições pobremente alinhadas e as regiões muito divergentes, através do programa TrimAl (versão 1.2) (parâmetros -gt 0.8 -st 0.001 -cons 60) (Capella-Gutiérrez *et al.*, 2009). Esses parâmetros removem todas as colunas com “gaps” em mais de 20% das sequências ou com um “score” de similaridade abaixo de 0.001, a não ser que que isso remova mais de 40% das colunas do alinhamento original, pois o objetivo é conservar pelo menos 60% deles. A partir do arquivo de saída (alinhamento) gerado pelo TrimAl este foi convertido em formato PHYLIP através de um “script” em “perl” e árvores filogenéticas foram construídas com o programa RaxML-HPC (“Randomized Axelerated Maximum Likelihood”) v 7.4.2 (Stamatakis, 2006) com as opções PROTGAMMAWAG e 100 réplicas de “bootstrap”. A visualização e edição das árvores foram realizadas através do programa MEGA 5.1 (Tamura *et al.*, 2011). O modelo evolutivo foi determinado pelo programa Modelgenerator (versão 0.82) (Keane *et al.*, 2006), antes da execução do RaxML-HPC. Ao final, os resultados dos alinhamentos múltiplos foram visualizados pelo programa Jalview versão 2 (Waterhouse *et al.*, 2009)

para a verificação manual dos motivos conservados específicos de  $\beta$ -lactamases (conforme lustrado na Figura 4.3).

Para a obtenção de sequências representativas foi utilizada uma análise de BLASTP (versão 2.2.21) das sequências obtidas nos itens 4.7.2 e 4.10 contra a base Refseq de proteínas (2780 Mb, março de 2012), utilizando um valor mais conservador de “E-value”, de  $10^{-50}$ . Para obter uma maior confiança na anotação da função dessas sequências representativas foi realizada uma análise com RPS-BLAST destas contra a base de domínios conservados CDD, com um “E-value” de  $10^{-20}$  como ponto de corte.

Para a classe A uma abordagem a mais foi utilizada como restrição, pois muitas sequências (89.568) foram recuperadas a partir do “hmmsearch” contra a base CAMERA (conforme descrito posteriormente no item 5.6) e muitas apresentaram um alto grau de similaridade (>95% de similaridade). Sendo assim, as sequências resultantes da busca pelo “hmmsearch” foram submetidas ao programa CD-HIT com a opção de 97% (“-n 0.97”) a fim de obter as sequências que apresentem no máximo 97% de similaridade entre si e retirar da análise aquelas muito similares, permanecendo aquelas que possuem maior variabilidade. Como a árvore ficou muito extensa e de difícil visualização e interpretação devido, ainda, ao elevado número de sequências, foram geradas duas árvores filogenéticas: uma contendo as prováveis  $\beta$ -lactamases mineradas da base de esgotos hospitalares e suas respectivas sequências representativas e a outra contendo as prováveis  $\beta$ -lactamases mineradas das bases CAMERA e IMG/M e suas respectivas sequências representativas.

Abaixo segue um esquema resumido da metodologia “in silico” descrita.



**Figura 4.3 Fluxograma resumido da metodologia referente à mineração e análise da diversidade de serina  $\beta$ -lactamases dispostas em projetos metagenômicos e esgotos hospitalares do Rio de Janeiro (itens 4.6 ao 4.10 e 4.12).**

#### 4.13 Caracterização dos metagenomas hospitalares do Rio de Janeiro

A exploração da diversidade taxonômica e funcional da comunidade microbiana presente nos esgotos hospitalares foi realizada a partir de resultados obtidos do alinhamento local, com o programa BLAST, das sequências de esgotos hospitalares contra (i) a base de dados de nucleotídeos não redundantes (NT, 41 Gb, 15.885.136 sequências, versão março de 2012), do NCBI, pelo programa BLASTN (Zhang *et al.*, 2000), e (ii) a base de dados



de proteínas não redundantes (NR, 13 Gb, 21.622.202 sequências, versão de janeiro de 2013), do NCBI, pelo programa BLASTX (Altschul *et al.*, 1997), respectivamente. Os resultados do BLAST foram visualizados e analisados pelo programa MEGAN 4 (“Metagenome Analyzer”) (Huson *et al.*, 2007), conforme descrito no item 4.13.1. Este programa utiliza como arquivo de entrada os resultados de busca pelo BLAST de um conjunto de dados de sequências metagenômicas (“query”) contra uma ou mais bases de dados (NT, NR, ENV-NT, ENV-NR, genomas). A partir daí são gerados gráficos de distribuição taxonômica das amostras a partir da atribuição das sequências a nós diferentes na taxonomia do NCBI (NCBI Taxonomy), utilizando o algoritmo LCA (“Last Common Ancestor” ou último ancestral comum) que se baseia no último ancestral comum.

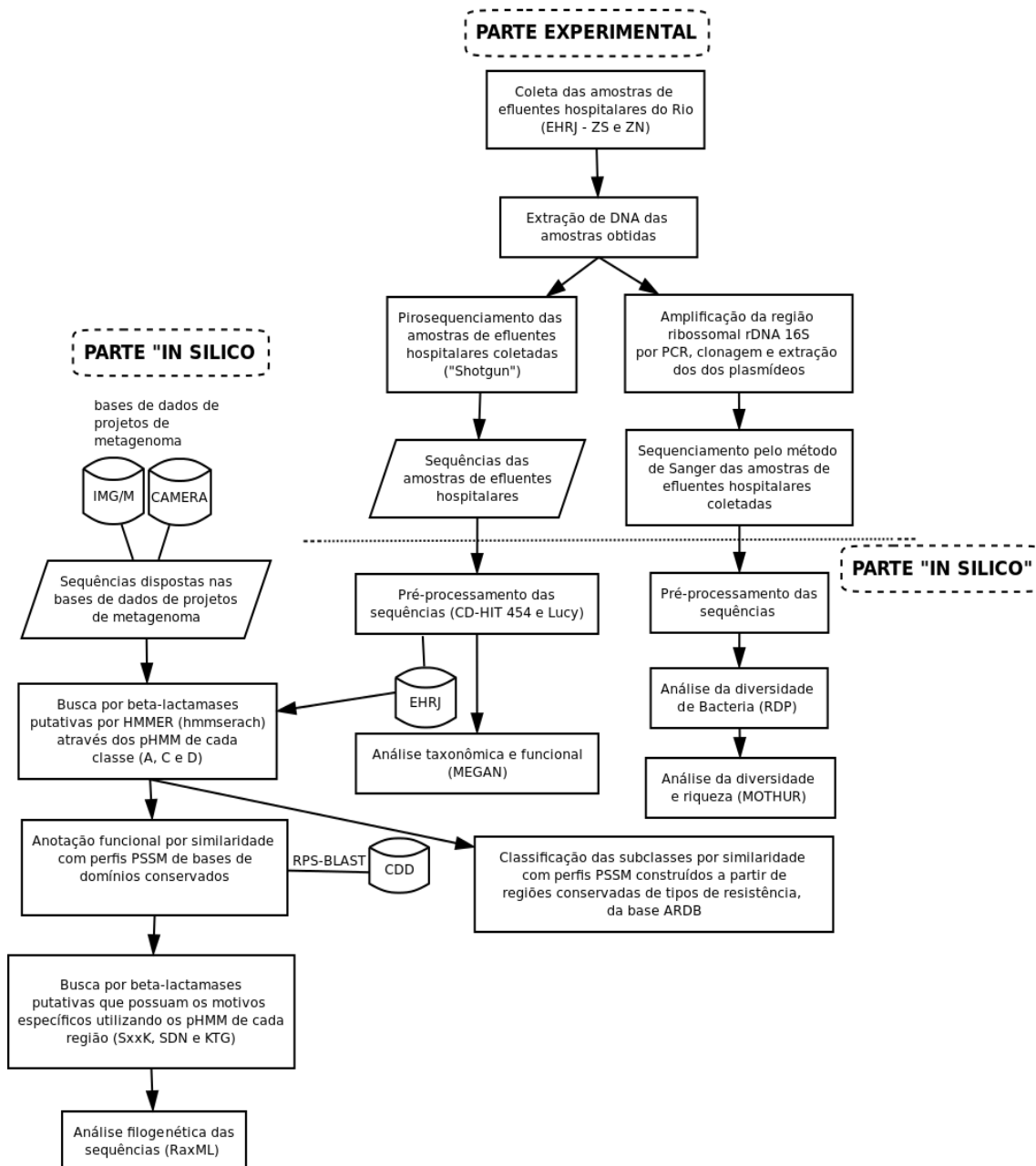
No algoritmo de LCA o nó ancestral mais recente (último ancestral comum), que é compartilhado pelos 5 melhores “hits”, é considerado a melhor classificação taxonômica para aquela sequência.

A classificação funcional do MEGAN, a partir do SEED é baseada na sequência de proteína com função que apresenta maior valor do score no BLAST (Huson *et al.*, 2007). Este programa é capaz de analisar grandes conjuntos de dados (sequências) sem a necessidade da montagem e é capaz de prover saídas gráficas e estatísticas para a comparação de diferentes grupos de dados. Na Figura 4.4 pode ser observado um esquema geral do funcionamento do MEGAN 4.

#### **4.14 . Curva de rarefação**

A curva de rarefação foi calculada e gerada pelo programa MEGAN 4, a partir dos dados do pirosequenciamento (item 4.5.5), em nível de gênero. Essa curva mostra os números de sequências de Bacteria, atribuídas de acordo com a classificação do NCBI Taxonomy, que apresentaram “hit” a partir dos resultados de BLAST e foi classificado dentro de algum gênero, de acordo com o algoritmo do último ancestral comum, o LCA.

O esquema a seguir resume de forma geral a metodologia empregada neste trabalho (Figura 4.4).



**Figura 4.4 Fluxograma resumido da metodologia descrita neste trabalho, incluindo a parte experimental e a parte "In silico", que envolve também as análises da parte experimental.**

## 5 RESULTADOS

### PARTE I: Parte experimental

#### 5.1 Análise físico-química e contagem de coliformes totais e fecais das amostras de EHRJ

A Tabela 5.1 apresenta os dados resultantes das análises de pH, DQO, DBO<sub>5</sub>, fósforo (P) e nitrogênio (N) das amostras dos dois hospitais. O valor de pH foi bem similar, assim como o valor de teores de N. Porém, os valores de DBO e DQO foram mais elevados na amostra do hospital da Zona Sul (ZS), enquanto que os valores de teores de P foram mais elevados na amostra do hospital da Zona Norte (ZN).

**Tabela 5.1 Análise físico-química das amostras de esgotos de dois hospitais situados na Zona Sul e Zona Norte da cidade do Rio de Janeiro.**

Parâmetros	Hospital Zona Sul (ZS)	Hospital Zona Norte (ZN)
pH	7,3	7,2
DQO (mg/L)	632	397
DBO (mg/L)	295	209
P (mg/L)	12,2	14,0
N (mg/L)	55,3	55,0

#### 5.2 Extração e quantificação do DNA metagenômico dos EHRJ

A extração e quantificação do DNA metagenômico das amostras revelou que estas não possuíam rastros e estavam em quantidade satisfatória (~2µg/µL) para o sequenciamento na plataforma 454/Life Science.

### **5.3 Caracterização do DNA metagenômico dos EHRJ**

#### **5.3.1 Amplificação por PCR da região do 16S ribossomal, clonagem da biblioteca e sequenciamento pelo método de Sanger**

A amplificação do 16S ribossomal (rrs) de Bacteria por PCR a partir do DNA extraído das amostras de esgotos hospitalares (ZN e ZS) (descrito nos itens 4.4 e 4.5.1) foi bem sucedida, resultando em bandas de intensidade moderada (~5ng de DNA). A clonagem da biblioteca ribossomal de Bacteria gerou, aproximadamente, 347 clones de 16S ribossoma (rrs) I de cada amostra (ZN e ZS). O sequenciamento do 16S ribossomal, após a extração plasmidial, resultou em sequências de tamanho aproximado de 500 pb. Após o pré-processamento das sequências geradas (conforme descrito no item 4.5.3) restaram 347 e 262 sequências de alta qualidade ( $Q > 20$ ) das amostras ZN e ZS, respectivamente. Além disso, foram retiradas 21 sequências quiméricas das amostras detectadas com o programa Decipher (Wright *et al.*, 2011).

#### **5.3.2 Estimativa da diversidade e riqueza de UTOs**

Na Tabela 5.2 estão apresentados os índices de Chao e ACE (para riqueza) e de Shannon (para diversidade), e a estimativa do número de UTOs, considerando uma distância evolutiva de 0,03. Pode-se observar que o número de UTOs (186) foi bem mais elevado no hospital ZN do que no ZS, mostrando uma riqueza maior nessa amostra. Na amostra ZS apenas 73 UTOs representaram a riqueza dos 247 clones obtidos, indicando um menor número de espécies presentes na amostra no momento da coleta. Comparando os valores dos índices de diversidade Chao e ACE entre as amostras ZS e ZN, se observou novamente que a amostra ZN apresentou uma maior riqueza microbiana. A comparação do valor do índice de Shannon entre as amostras revelou que a diversidade do hospital ZN é maior que o ZS, corroborando com os números de UTOs obtidas.

**Tabela 5.2 Estimativa de riqueza de UTOs e índices de diversidade calculados a partir de bibliotecas de 16S ribossomal de Bacteria de dois hospitais públicos do Rio de Janeiro.**

<sup>a</sup> Amostra	Biblioteca	<sup>b</sup> NS	<sup>c</sup> Nº UTOs	Chao1	ACE	Shannon
ZS	DNA 16S	297	73	171 (117,3; 291*)	301,4 (221,4; 424,6*)	3,0 (2,8; 3,2*)
ZN	DNA 16S	270	186	1118 (706; 1853*)	2.218 (1.711; 2.894*)	4,9 (4,7; 5,01*)

<sup>a</sup> Amostras coletadas em dois hospitais públicos distintos localizados na cidade do Rio de Janeiro. ZS: Hospital localizado na Zona Sul. ZN: Hospital localizado na Zona Norte do Rio.

<sup>b</sup> Número de sequências de cada biblioteca.

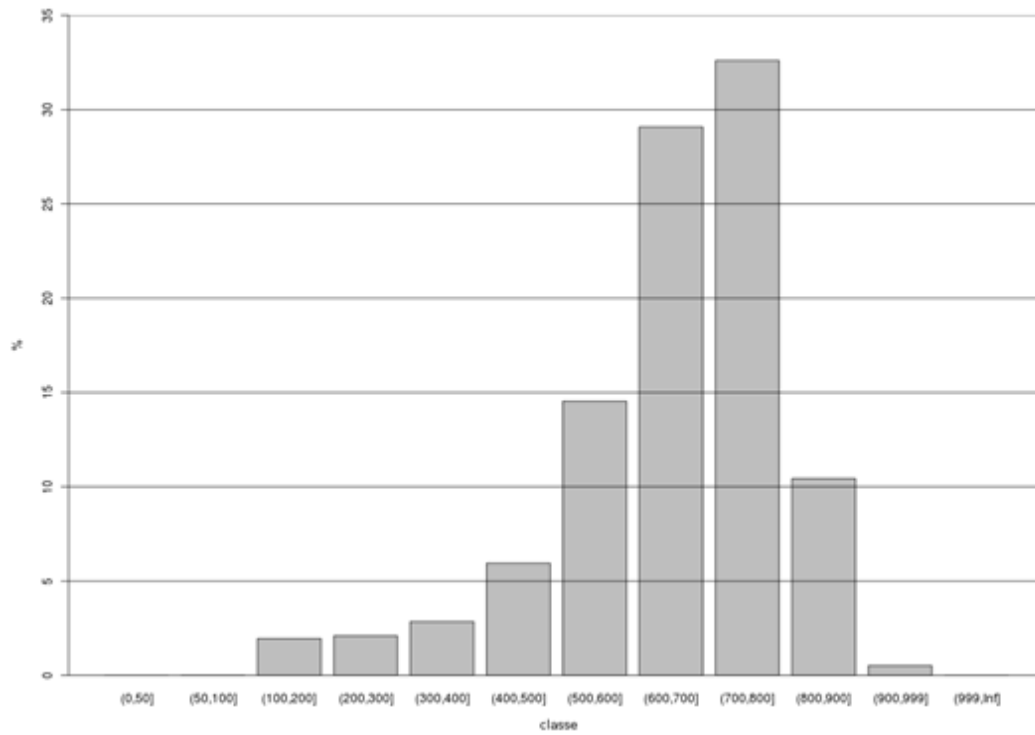
<sup>c</sup> Número de UTOs. Calculado com o programa MOTHUR com 3% de distância.

\*Os números entre parênteses representam os intervalos de confiança dos índices;

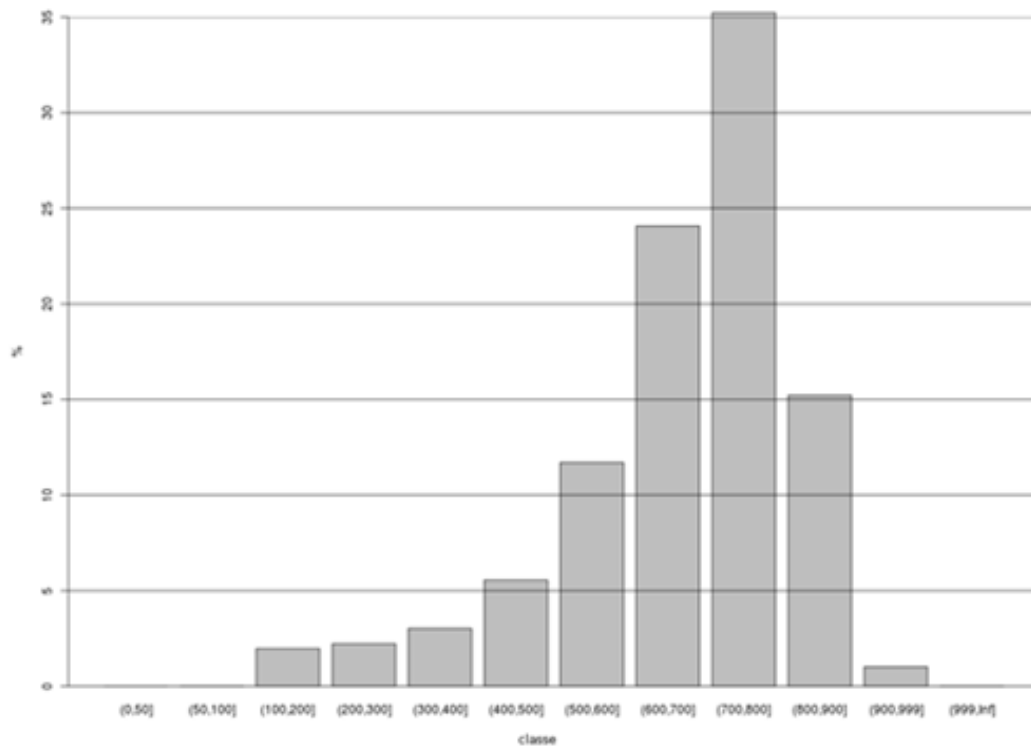
### **5.3.3 Informações sobre os dados gerados pelo pirosequenciamento das amostras de EHRJ**

Na primeira corrida do pirosequenciamento foram geradas, após a etapa do filtro, 476.821 sequências para a réplica 01 do hospital ZS, com uma média de 652pb, enquanto para a réplica 02 do mesmo hospital foram geradas 437.354 sequências com uma média de 668 (Tabela 5.3). A maioria das sequências continha entre 700-800 pares de base, conforme indicado no histograma das Figuras 5.1 (A) e (B). Antes da etapa do filtro havia 565.344 e 682.225 sequências para as réplicas 01 e 02, respectivamente.

(A)



(B)



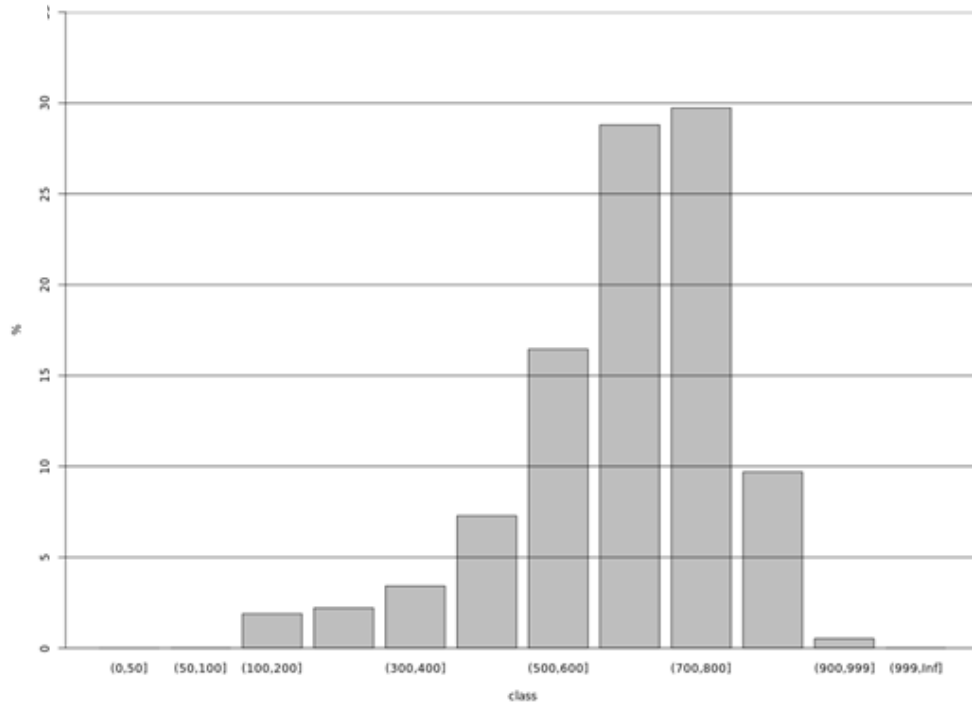
**Figura 5.1** Distribuição do comprimento das seqüências do pirosequenciamento das réplicas 01 (A) e 02 (B) coletadas em um hospital público da Zona Sul do Rio de Janeiro (ZS), após serem filtradas por qualidade ( $Q > 20$ ), redundância e tamanho ( $> 100$  nucleotídeos).

**Tabela 5.3 Dados numéricos das sequências das réplicas 01 (ZS01) e 02 (ZS02) do hospital da Zona Sul do Rio de Janeiro, geradas após o pirosequenciamento com a plataforma 454/Life Science e filtradas por qualidade (Q> 20), redundância e tamanho (>100 nucleotídeos).**

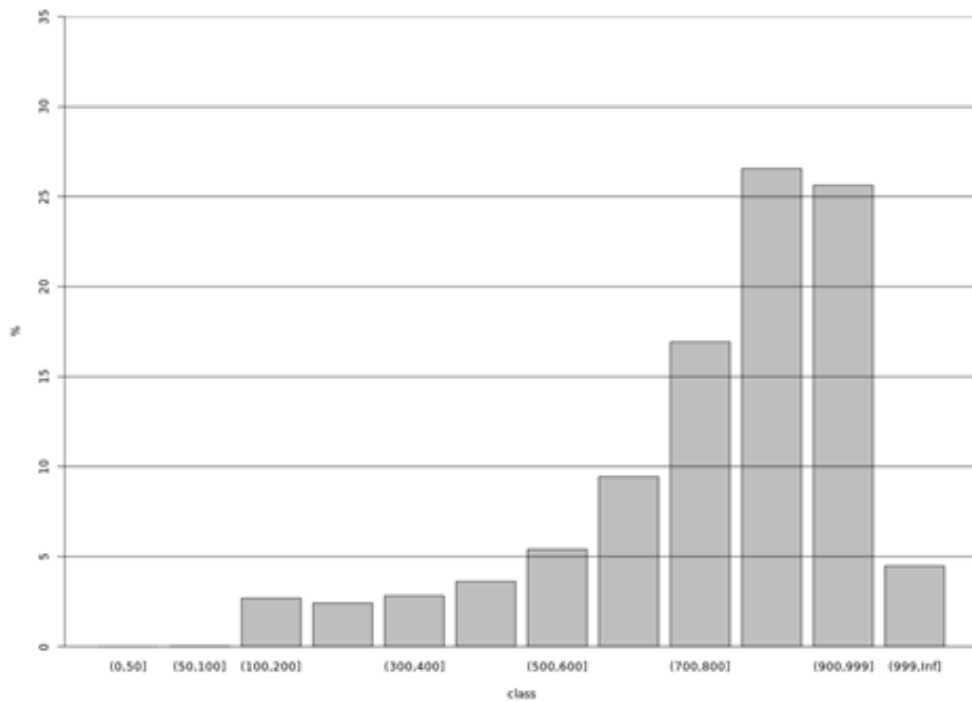
	ZS01	ZS02
Nº total de sequências	476	437
Menor tamanho de sequência	100	100
Maior tamanho de sequência	1.372	1.533
Média	652	668
Mediana	682	705
Moda	724	755

Na segunda corrida, após a etapa do filtro, foram geradas 699.252 sequências para réplica 01 da amostra de efluente do hospital ZN, com uma média de 752 pb, enquanto para a réplica 02 foram geradas 618.486 sequências, com uma média de 771 pb (Tabela 5.4). A maioria das sequências resultantes do pirosequenciamento da réplica 01 da amostra ZN apresentou tamanhos entre 800-900 pares de base, conforme indicado no histograma das Figuras 5.2 (A) e (B). Antes da etapa do filtro havia 774.961 e 682.225 sequências para as réplicas 01 e 02, respectivamente.

(A)



(B)



**Figura 5.2** Distribuição do comprimento das seqüências do pirosequenciamento das réplicas 01 (A) e 02 (B) coletadas em um hospital público da Zona Norte do Rio de Janeiro (ZN), após serem filtradas por qualidade ( $Q > 20$ ), redundância e tamanho ( $> 100$  aminoácidos).



**Tabela 5.4 Dados numéricos das sequências das réplicas 01 (ZN01) e 02 (ZN02) do hospital da Zona Norte do Rio de Janeiro, geradas após o pirosequenciamento com a plataforma 454/Life Science e filtradas por qualidade (Q> 20), redundância e tamanho (>100 nucleotídeos).**

	<b>ZN01</b>	<b>ZN02</b>
Nº total de sequências	775	682
Menor tamanho de sequência	100	100
Maior tamanho de sequência	1.775	1.765
Média	752	771
Mediana	820	828
Moda	919	919

No total foram geradas 914.175 sequências referentes às amostras coletadas no hospital da Zona Sul (ZS) e 1.317.738 sequências referentes às amostras coletas no hospital da Zona Norte (ZN).

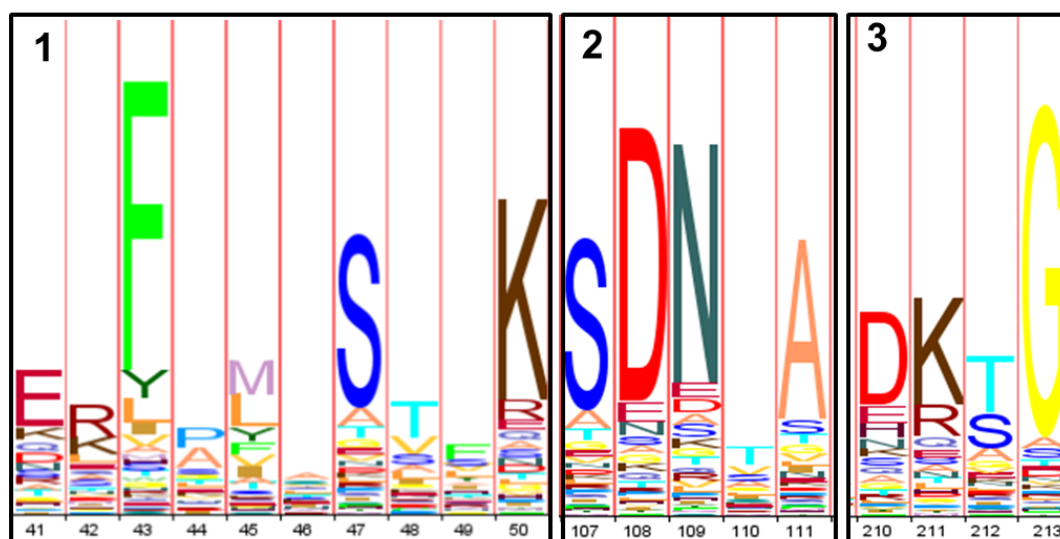
## **PARTE 2: “In silico” (Não experimental)**

### **5.4 Pré-processamento das sequências geradas pela plataforma 454/Life Science**

Em relação ao resultado do pirosequenciamento das amostras do hospital ZS do Rio, após a filtragem utilizando os programas CD-HIT 454 e Lucy (item 4.5.6) foram obtidas 476.821 e 437.354 sequências relacionadas às réplicas 01 e 02, respectivamente (Tabela 5.3). A filtragem removeu apenas 18% das sequências obtidas, considerando as duas réplicas juntas como uma amostra. Já para as amostras hospitalares ZN foram geradas 699.252 e 618.486 sequências relacionadas às réplicas 01 e 02, respectivamente (Tabela 5.4). A filtragem removeu apenas 10% das sequências obtidas, considerando a soma das réplicas 01 e 02 da amostra ZN.

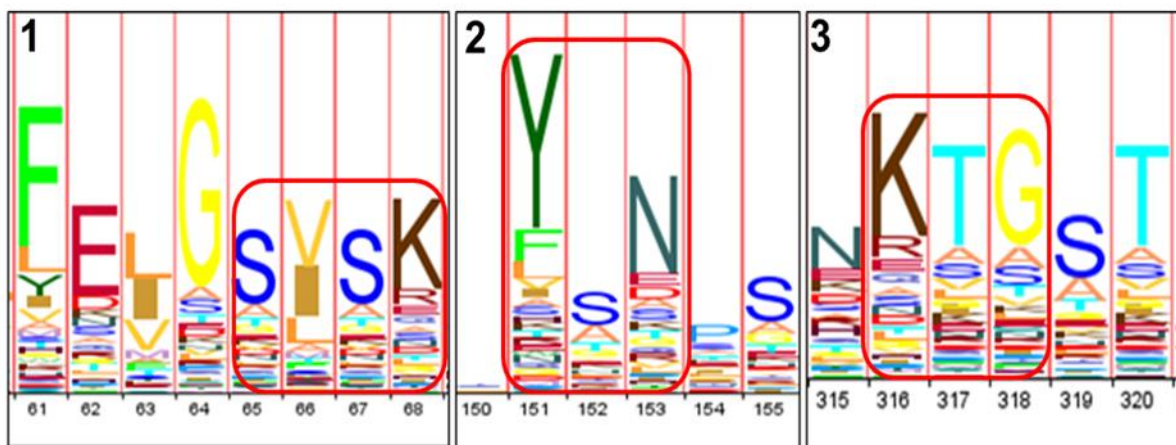
## 5.5 Construção de pHMMs de serina $\beta$ -lactamases

A partir da visualização dos logos dos pHMMs (Schuster-Boeckler *et al.*, 2004) construídos para cada classe de  $\beta$ -lactamases, baseado no alinhamento “semente”, estes se mostraram adequados para a busca por homólogos com o “hmmsearch” do pacote HMMER. As Figuras a seguir (5.3, 5.4 e 5.5) apresentam as principais regiões conservadas dos logos de cada perfil. No logo da classe A, janela 1 da Figura 5.3, está ilustrado o sítio catalítico Ser-X-X-Lys [S-x-x-K], nas posições 47-50, juntamente com o aminoácido Phe (F) que antecede o sítio catalítico (na posição 43), formando uma região conservada, encontrado na classe A. Os aminoácidos (x-x) que compõem o sítio catalítico, entre Ser e Lys, variaram bastante e foram: (i) para a primeira posição a Treonina (T) e a Valina (V), em menor distribuição; (ii) e para a segunda posição após a Serina, as frequências foram baixas, sendo a Fenilalanina (F) o aminoácido mais frequente dentre todos. Antecedendo o sítio catalítico (por 4 sítios), o aminoácido Fenilalanina (F) foi o mais frequente. O segundo motivo (janela 2) foi o Ser-Asp-Asn [S-D-N] que também apareceu com maior frequência, sendo que a serina (S) apareceu com uma frequência menor que o aspartato (D) e a asparagina (N). Os aminoácidos que compõem o terceiro motivo (janela 3), a tríade Lys-Thr-Gly [K-T-G], também foram frequentes, como esperado para um bom perfil HMM de  $\beta$ -lactamases (Figura 5.3).



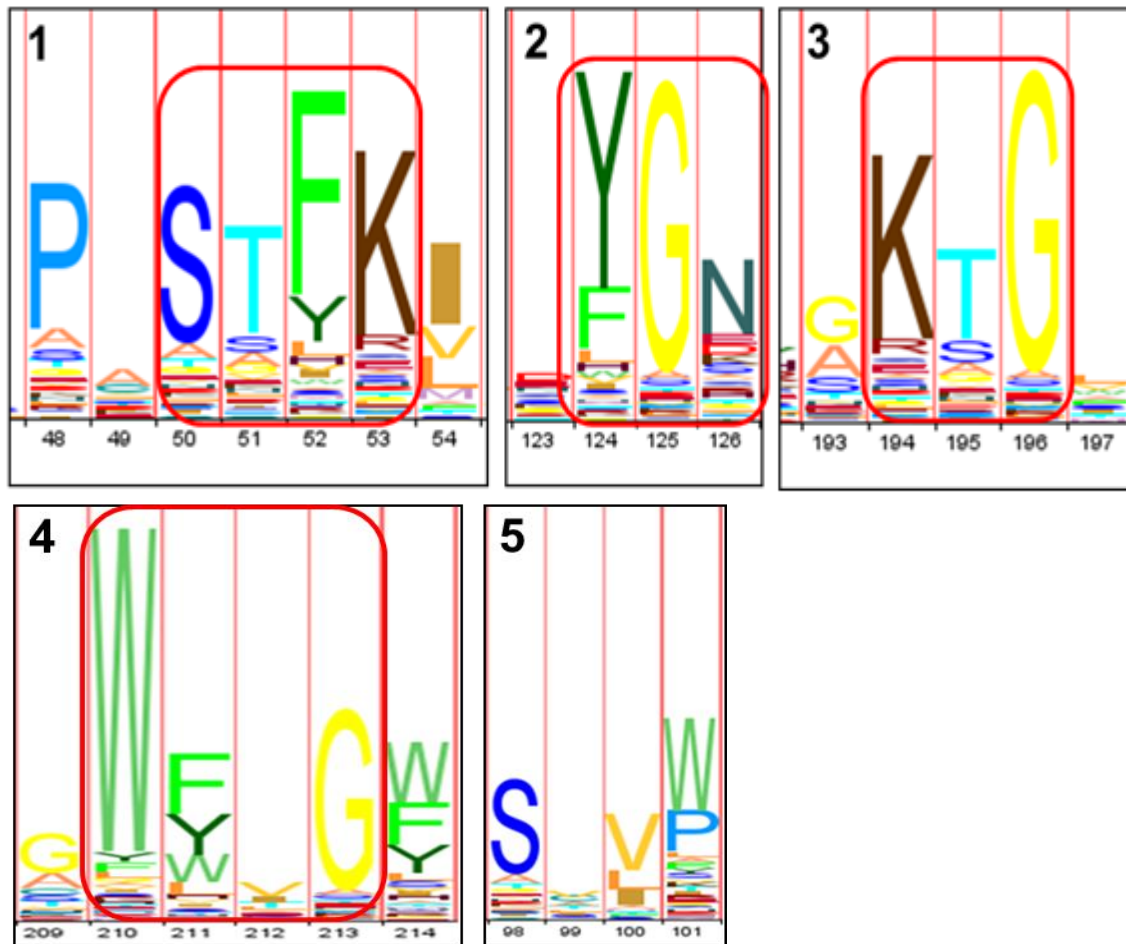
**Figura 5.3** Imagem ilustrativa do logo parcial do pHMM dos principais motivos da  $\beta$ -lactamase classe A, gerado pelo programa LogoMat-M que se encontram nas caixas 1,2 e 3. (1) Primeiro motivo: sítio catalítico; (2) Segundo motivo; (3) Terceiro motivo. Os números abaixo das letras indicam a posição de cada uma no logo.

O logo do perfil da classe C também se mostrou adequado, apresentando as principais regiões conservadas das serina  $\beta$ -lactamases. Em destaque, na Figura 5.4, estão três principais regiões conservadas em diferentes janelas (1, 2 e 3). A primeira janela do logo (1) mostra a região conservada F-E-L(I ou V)-G-S-V(I)-S-K, onde está localizado o sítio catalítico Ser-X-X-Lys [S-x-x-K]. Observou-se que o segundo aminoácido variável (x) é mais frequente como serina (S), característico de enzimas  $\beta$ -lactamases classe C. O segundo motivo (janela 2) que apareceu com maior frequência foi o Tyr-Ser(Ala)-Asn [Y-S(A)-N] e o terceiro motivo (janela 3) foi o Lys-Thr-Gly [K-T-G], seguido de mais dois aminoácidos conservados, a serina (S) e treonina (T), que também se apresentaram em maior frequência no logo (Figura 5.4).



**Figura 5.4** Imagem ilustrativa do logo parcial do pHMM dos principais motivos da  $\beta$ -lactamase classe C, gerado pelo programa LogoMat-M. que se encontram nas caixas 1,2 e 3. (1) Primeiro motivo: sítio catalítico; (2) Segundo motivo; (3) Terceiro motivo. Os números abaixo das letras indicam a posição de cada uma no logo.

O logo do perfil da classe D também se mostrou adequado, apresentando as principais regiões conservadas das  $\beta$ -lactamases pertencentes a essa classe. Para esse perfil destacam-se cinco principais regiões conservadas, ilustradas nas cinco janelas da Figura 5.5. A primeira janela do logo (1), ilustra a região conservada Pro-x-Ser-Thr-Phe(Tyr)-Lys-Ile(Val/Leu) [P-x-S-T-F(Y)-K-I(V/L)], onde está localizado o sítio catalítico Ser-X-X-Lys [S-x-x-K]. Observou-se que os aminoácidos treonina (T) e fenilalanina (F) presentes na região do sítio catalítico foram mais frequentes no primeiro e segundo aminoácidos variáveis (x-x), característico de enzimas classe D. O segundo motivo (janela 2) mais frequente foi Tyr(Phe)-Gly-Asn [Y(F)-G-N], o terceiro motivo (janela 3) foi o Lys-Thr(Ser)-Gly [K-T(S)-G], o quarto (janela 4) foi a téttrade Trp-x-x-Gly [W-X-X-G] e o quinto (janela 5) foi a tríade Ser-x-Val (Leu) [S-X-V (L)] (Figura 5.5).



**Figura 5.5** Imagem ilustrativa do logo parcial do pHMM dos principais motivos da  $\beta$ -lactamase classe D, gerado pelo programa LogoMat-M. que se encontram nas caixas 1,2 e 3. (1) Primeiro motivo: sítio catalítico; (2) Segundo motivo; (3) Terceiro motivo; (4) Quarto motivo; (5) Quinto motivo. Os números abaixo das letras indicam a posição de cada uma no logo.

## 5.6 Busca por homólogos de serina $\beta$ -lactamases

A busca por homólogos de  $\beta$ -lactamases com os p HMMs contra a base de dados ambientais CAMERA e IMG/M retornaram 52.122 e 2.234 “hits”, respectivamente, para a classe A, 2.384 e 10.420 “hits” para a classe C e 804 e 1.894 “hits” para a classe D. Os resultados referentes à busca do “hmmsearch” contra a base de esgotos hospitalares retornaram os seguintes valores para os hospitais ZS e ZN, respectivamente: 138 e 149 para classe A, 41 e 57 para a classe C e 40 e 185 para a classe D (Tabela 5.5).

**Tabela 5.5 Resultado dos números de sequências de prováveis homólogos de  $\beta$ -lactamases obtidas após a busca com o “hmmsearch” com os perfis de cada classe (pré-filtro), e após o “hmmsearch” utilizando os perfis de regiões específicas de  $\beta$ -lactamases (pós-filtro) contra as bases CAMERA e IMG/M e, a porcentagem de sequências que apresentaram anotação de  $\beta$ -lactamase após o RPS-BLAST contra a base COG para as classes A e D e Pfam para a classe C, pré-filtro e pós-filtro.**

<b>Classe de <math>\beta</math>-lactamase/ Base de dados</b>	<b>*Nº de sequências recuperadas pré-filtro</b>	<b>% de sequências putativas de <math>\beta</math>-lactamase (pré-filtro)</b>	<b>Nº de sequências recuperadas pós-filtro</b>	<b>% de sequências putativas de <math>\beta</math>-lactamase (pós-filtro)</b>
A/ CAMERA	52.122	99,7	2126	4,07
A/ IMG_M	2.234	95	65	2,76
C/ CAMERA	2.384	99,6	27	1,13
C/ IMG_M	10.420	99,2	125	1,19
D/ CAMERA	804	27	22	0,74
D/ IMG_M	1.894	42	55	1,22

\* Após o “hmmsearch” foi utilizado um filtro (“script” na linguagem ruby) para remover sequências menores que 100 aminoácidos e o CD-HIT 100 para retirar as sequências redundantes (com 100% de identidade). Os valores apresentados na Tabela acima são após esse processo.

### **5.7 Anotação funcional das sequências homólogas às serina $\beta$ -lactamases**

Os resultados da análise por similaridade com bases de domínios conservados que apresentaram maiores valores (“hits”) foram contra a base COG (classe de  $\beta$ -lactamases A e D) e Pfam (classe C). A partir desses resultados de similaridade através do RPS-BLAST (item 4.8), verificou-se que a maioria das sequências homólogas à classe A (99,7% nos resultados do CAMERA e 95% nos resultados do IMG) e C (99,6 e 99,2% para o CAMERA e IMG, respectivamente) apresentaram anotação de  $\beta$ -lactamases. Já para a classe D apenas 27 e 42% das sequências resultantes do “hmmsearch”

apresentaram anotação com função de  $\beta$ -lactamase e/ou  $\beta$ -lactamase classe D.

Com o intuito de caracterizar apenas aquelas sequências que realmente possuíam os motivos principais de serina  $\beta$ -lactamases, uma nova busca com o HMMER foi realizada, agora com os perfis de cada motivo específico de cada classe, conforme descrito no item 4.14. A classe A ficou com 2.126 “hits” (CAMERA) e 338 “hits” (IMG), a classe C apresentou 27 “hits” (CAMERA) e 125 (IMG) sequências e a classe D apresentou 22 (CAMERA) e 55 (IMG). Um procedimento similar foi realizado com as sequências oriundas dos esgotos hospitalares. Nessas amostras esta avaliação foi realizada manualmente para a presença dos sítios uma vez que as sequências não estavam completas. Dos resultados obtidos pelo “hmmsearch” para as três classes (A, C e D) de prováveis  $\beta$ -lactamases, mais de 85% delas apresentaram anotação de  $\beta$ -lactamase a partir dos resultados do RPS-BLAST. A Tabela 5.5 mostra a quantidade de prováveis homólogos de  $\beta$ -lactamases obtidas após a busca pelo “hmmsearch” usando o primeiro pHMM (conforme descrito no item 4.6) de cada classe (pré-filtro) e depois, após o “hmmsearch”, utilizando o perfil de cada região específica para cada classe contra as bases de dados CAMERA e IMG/M (pós-filtro).

Em relação às amostras de esgotos hospitalares do Rio de Janeiro (EHRJ), a classificação da busca por similaridade a partir da base PSSM de domínio conservado COG apresentou 91% das sequências de ZS recuperadas com anotação de  $\beta$ -lactamase classe A, enquanto ZN apresentou 96,7%. Considerando a classe C, 86% das sequências de ZS apresentaram anotação de  $\beta$ -lactamase, enquanto 95% das sequências de ZN apresentaram a mesma classificação. Considerando a classe D, 97,4% das sequências de ZS e 94% de ZN apresentaram anotação de  $\beta$ -lactamase classe D. Esses resultados podem ser visualizados de forma mais organizada na Tabela 5.6. Apenas as sequências que apresentaram anotação de  $\beta$ -lactamase e possuíam as regiões conservadas relacionadas a cada classe foram utilizadas para a análise filogenética.

Como o resultado relacionado às prováveis  $\beta$ -lactamases classe A apresentou um elevado número de sequências ainda após a busca do “hmmsearch” usando os perfis das regiões específicas contra a base CAMERA (2126 sequências de aminoácidos) e estas eram muito similares (> 95%

similaridade), utilizou-se o CD-HIT V4.5.7 (Li *et al.*, 2001) com a opção de 97% de similaridade (“-n 0.97”) para retirar as sequências muito similares, conforme descrito no item 4.12. Após o CD-HIT restaram 56 sequências de aminoácidos prováveis homólogos da classe A do CAMERA e 50 do IMG/M, mais as 42 e 22 sequências das réplicas 01 e 02 das amostras ZS (ZS01 e ZS02) e 22 e 39 sequências das réplicas 01 e 02 das amostras ZN (ZN01 e ZN02), respectivamente. Muitas sequências prováveis de  $\beta$ -lactamase oriundas da base de efluente hospitalar, recuperadas pelo HMMER (“hmmsearch”), não tinham o tamanho integral da enzima (faltava, principalmente, a região K-T-G). Sendo assim, as árvores filogenéticas foram construídas com estas sequências parciais já que elas apresentavam mais de 140 aminoácidos, com o tamanho variando entre 140-340 e a anotação, após análise com RPS-BLAST contra o COG e Pfam, apresentou função de  $\beta$ -lactamase específica de cada classe. As classes do COG e Pfam obtidas foram: COG2367-“Beta-lactamase class A [Defense mechanisms]” (para a classe A), Pfam PF00144 - Beta-lactamase (para a classe C) e COG2602- “Beta-lactamase class D [Defense mechanisms]” (para a classe D). Os menores valores de “E-value” das sequências utilizadas para a construção das árvores foram  $10^{-42}$ ,  $10^{-180}$  e  $10^{-72}$  para as classes A, C e D, respectivamente. Já os maiores valores de “E-value” foram  $10^{-21}$ ,  $10^{-21}$  e  $10^{-25}$  para as classes A, C e D, respectivamente. A Tabela contendo tais resultados encontra-se no material suplementar disponibilizado em CD, anexado ao final do corpo da tese.



**Tabela 5.6 Resultado dos números de sequências obtidas após a busca com o “hmmsearch” (pré-filtro) a partir dos perfis de cada classe de  $\beta$ -lactamases (A, C e D) contra as bases de esgotos ZS e ZN. ZS: Hospital da Zona Sul. ZN: Hospital da Zona Norte.**

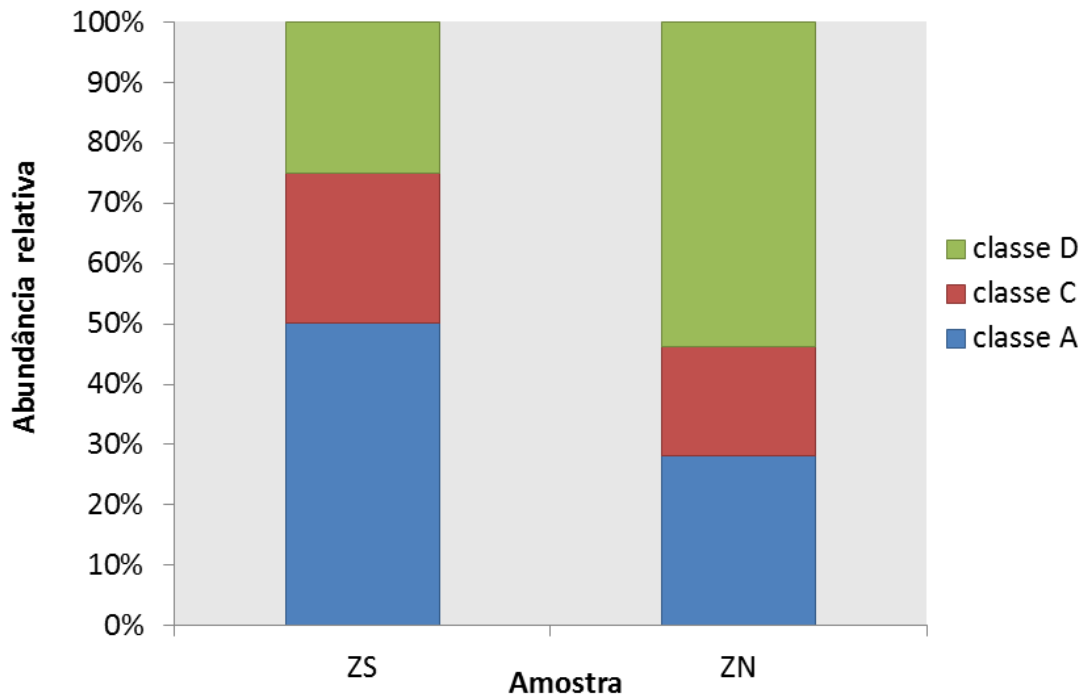
$\beta$ -lactamase/ Base de dados	Nº de sequências obtidas pré-filtro	% de sequências prováveis de $\beta$ -lactamase (pré-filtro)
Classe A/ ZS01 e ZS02	62 e 60	96 e 84
Classe A/ ZN01 e ZN02	22 e 39	95 e 98
Classe C/ ZS01 e ZS02	23 e 18	74* e 100
Classe C/ ZN01 e ZN02	26 e 31	92 e 97
Classe D/ ZS01 e ZS02	20 e 18	95,4 e 100
Classe D/ ZN01 e ZN02	84 e 78	94,5 e 93,3

\*O valor dessa porcentagem foi mais baixo porque os 26% restantes apresentaram anotação de proteínas hipotéticas. Não significa que não tenham função de  $\beta$ -lactamases.

## 5.8 Abundância relativa de $\beta$ -lactamases

### 5.8.1 *Em amostras de esgotos hospitalares*

Em relação às amostras de esgotos hospitalares ZS e ZN se observou uma maior ocorrência da classe A em ZS, com aproximadamente 50%, enquanto as classes C e D apresentaram uma frequência similar, com aproximadamente 25%. Já na amostra hospitalar ZN a classe D apresentou uma frequência maior, com aproximadamente 55%, enquanto as classes A e C também apresentaram frequências similares, com aproximadamente 25% e 20%, respectivamente (Figura 5.6).



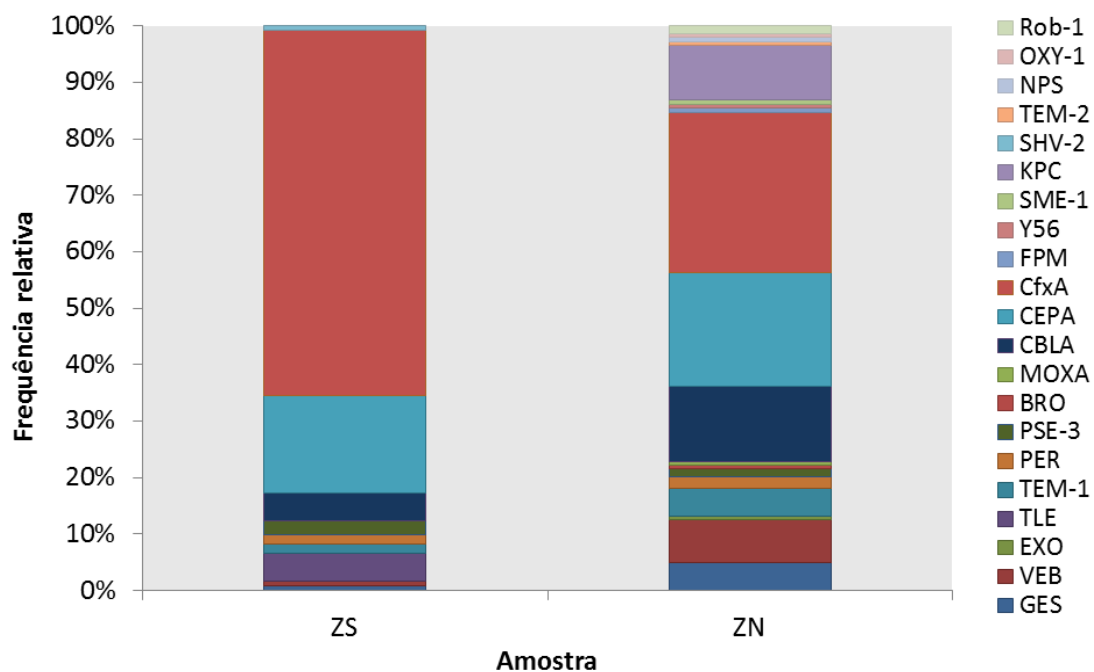
**Figura 5.6.** Abundância relativa (%) dos prováveis homólogos das três classes de serina  $\beta$ -lactamases relacionadas às amostras de esgotos hospitalares ZS (Zona Sul, réplicas 01 e 02) e ZN (Zona Norte, réplicas 01 e 02). As réplicas de cada amostra foram unidas.

#### **5.8.1.1 Abundância relativa das subclasses nos esgotos hospitalares do Rio de Janeiro**

A análise por similaridade com regiões conservadas utilizando perfis PSSM de genes de resistência a antibióticos dispostos no ARDB resultou na anotação de subclasses de  $\beta$ -lactamases classe A de 96% das sequências submetidas para as amostras ZS (132/138) e 94% para ZN (140/149). Em relação à classe C os resultados foram de 93% (38/41) e 100% (58/58) para ZS e ZN, respectivamente, e para a classe D foram classificadas 88% (143/163) e 90% (36/40) para ZS e ZN.

Em relação às subclasses de  $\beta$ -lactamases classe A presentes nas amostras de esgotos hospitalares ZS e ZN, se observou uma abundância maior da subclasse CfxA em ambas as amostras sendo que ZS apresentou mais da metade das sequências com essa classificação (65%) enquanto a ZN apresentou, aproximadamente, 28,4%. Na amostra ZN as demais subclasses mais frequentes foram CEPA (20%) e CBLA (13%). A enzima  $\beta$ -lactamase tipo KPC não foi detectada na amostra ZS enquanto na ZN apresentou uma taxa de

10%. Já as subclasses GES e VEB apresentaram uma baixa frequência em ZS (0,8%), enquanto que em ZN apresentaram uma taxa de 5% e 8%, respectivamente (Figura 5.7 e Tabela 5.7).



**Figura 5.7** Frequência relativa das subclasses das  $\beta$ -lactamases classe A presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.

**Tabela 5.7. Valores da frequência relativa (%) das subclasses pertencentes a  $\beta$ -lactamase classe A presente nas amostras hospitalares da Zona Sul (ZS) e Zona Norte (ZN), classificadas a partir da base de dados ARDB.**

Subclasse	Frequência relativa (%)	
	ZS	ZN
GES	0,82	4,86
VEB	0,82	7,64
EXO	0,00	0,69
TLE	4,91	0,00
TEM-1	1,64	4,86
PER	1,64	2,08
PSE-3	2,46	1,39
BRO	0,00	0,69
MOXA	0,00	0,69
CBLA	4,92	13,19
CEPA	17,2	20,14
CfxA	64,75	28,47
FPM	0,00	0,69
Y56	0,00	0,69
SME-1	0,00	0,69
KPC	0,00	9,72
SHV-2	0,00	0,00
TEM-2	0,00	0,69
NPS	0,00	0,69
OXY-1	0,00	0,69
Rob-1	0,00	1,39

Em relação às  $\beta$ -lactamases classe C presentes nas amostras de esgotos hospitalares (ZS e ZN), se observou uma distribuição mais heterogênea das subclasses. No entanto, a amostra hospitalar ZS apresentou uma abundância maior em relação ao tipo ACC, representando metade das sequências obtidas classificadas como provável classe C (aproximadamente 53%) enquanto as subclasses mais frequentes na amostra ZN foram FOX e ACC com frequência de, aproximadamente, 21% e 19%, respectivamente (Figura 5.8 e Tabela 5.8).

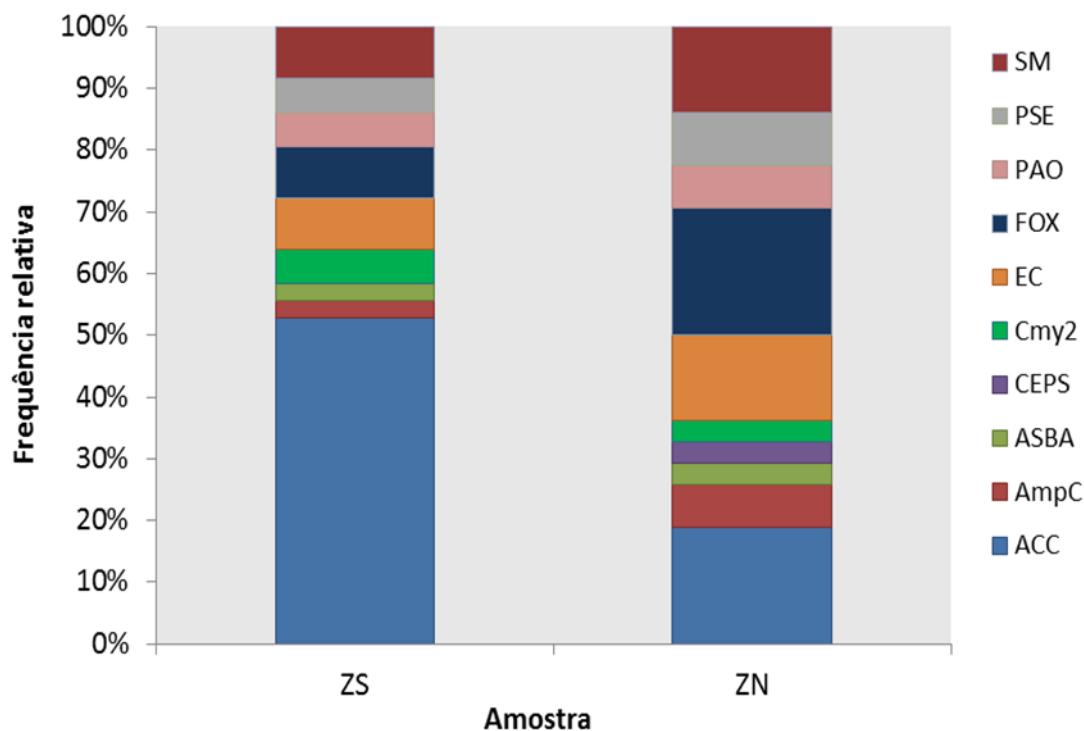
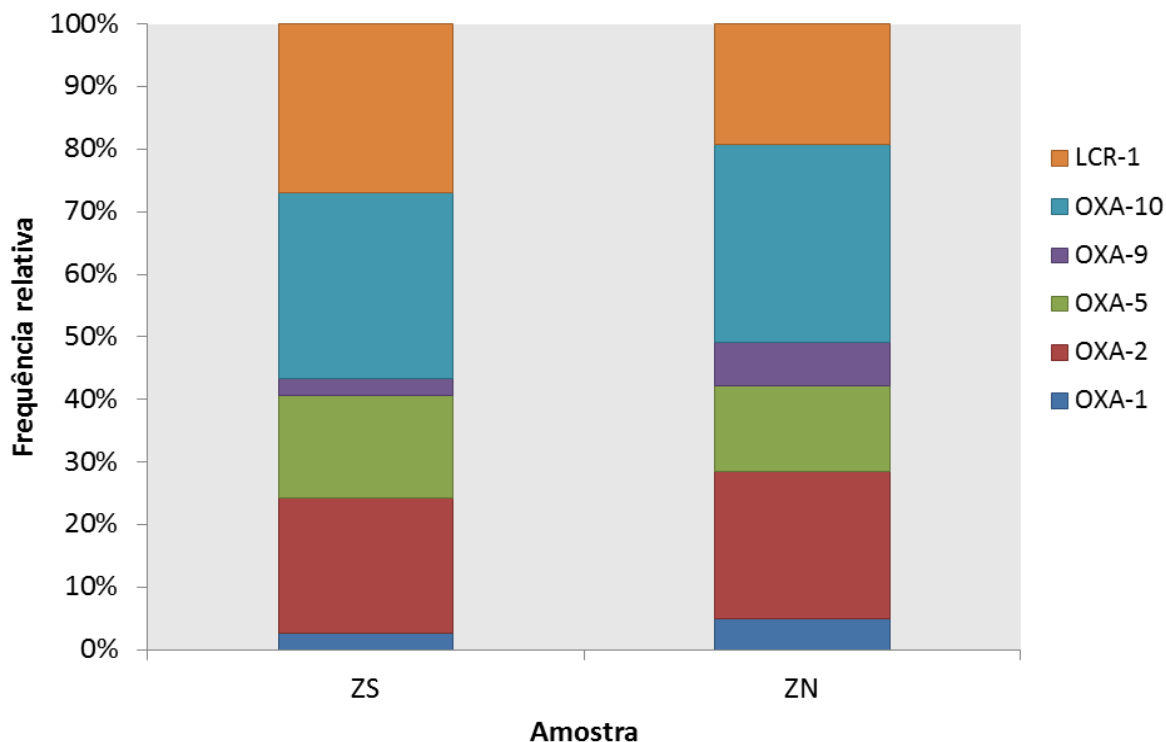


Figura 5.8. Frequência relativa (%) das subclasses das  $\beta$ -lactamases classe C presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.

Tabela 5.8. Valores da frequência relativa (%) das subclasses pertencentes a  $\beta$ -lactamase classe C presente nas amostras hospitalares da Zona Sul (ZS) e Zona Norte (ZN), classificadas a partir da base de dados ARDB.

Subclasse	Frequência relativa (%)	
	ZS	ZN
ACC	52,8	18,97
AmpC	2,78	6,9
ASBA	2,78	3,45
CEPS	0,00	3,45
CmY2	5,56	3,45
EC	8,33	13,79
FOX	8,33	20,69
PAO	5,56	6,9
PSE	5,56	8,62
SM	8,33	13,79

Considerando as  $\beta$ -lactamases classe D presentes nas amostras de esgotos hospitalares (ZS e ZN), se observou um perfil bem similar das subclasses. As mais frequentes, em ambas as amostras, foram as subclasses OXA-10 (30 e 32%), LCR-1 (27 e 19%) e OXA-2 (22 e 24%) para ZS e ZN, respectivamente (Figura 5.9 e Tabela 5.9).



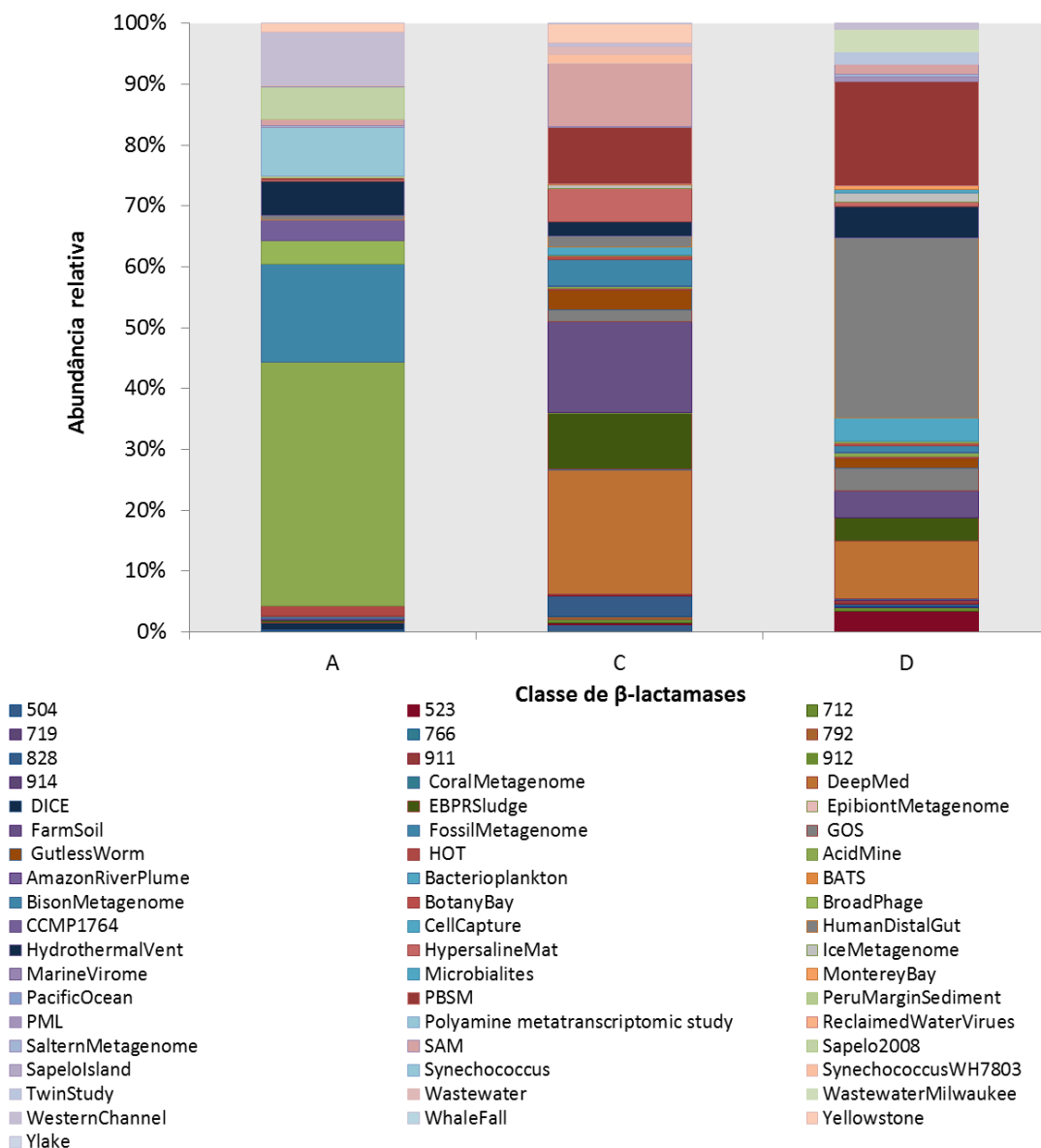
**Figura 5.9.** Frequência relativa (%) das subclasses da  $\beta$ -lactamase classe D presentes nas amostras de esgotos hospitalares ZS (Zona Sul) e ZN (Zona Norte) classificadas a partir da base de dados ARDB.

**Tabela 5.9.** Valores da frequência relativa (%) das subclasses pertencentes a  $\beta$ -lactamase classe D presente nas amostras hospitalares da Zona Sul (ZS) e Zona Norte (ZN), classificadas a partir da base de dados ARDB.

Subclasse	Frequência relativa (%)	
	ZS	ZN
OXA-1	2,7	4,97
OXA-2	21,62	23,6
OXA-5	16,2	13,66
OXA-9	2,7	6,83
OXA-10	29,73	31,68
LCR-1	27,03	19,25

### **5.8.2 Em projetos metagenômicos da base CAMERA**

Em relação à ocorrência dos prováveis homólogos de  $\beta$ -lactamases encontrados nos diferentes projetos de metagenoma da base CAMERA, se observou que os projetos “Acid Mine” (ambientes de mina ácida), “Bison Metagenome”, “Western Channel” e “Polyamine metatranscriptomic study” apresentaram uma grande quantidade de  $\beta$ -lactamases classe A, representando 40%, 17%, 9% e 8%, respectivamente, e quase nenhuma sequência das classes C e D. Já as  $\beta$ -lactamases classe C foram mais frequentes nos projetos “DeepMed”, “Farm Soil”, “SAM”, “EPBRSludge” e “PBSM” com 20%, 15%, 10%, 9% e 9%, respectivamente. As  $\beta$ -lactamases classe D foram mais abundantes nos projetos “Human Distal Gut” (Intestino médio de humanos), “PBSM” (Metagenoma das areias das praias do Pacífico) e “DeepMed” apresentando 30%, 17% e 9%, respectivamente, enquanto as classes A e C foram menos frequentes (Figura 5.10).



**Figura 5.10.** Frequência relativa (%) dos prováveis homólogos de serina  $\beta$ -lactamases classe A, C e D, minerados a partir dos perfis HMM, presentes nos diferentes projetos presentes na base CAMERA. DICE: “Dauphin Island Cubitainer Experiment”; Bison: “Metagenome from Yellowstone Bison Hot Spring”; HOT: Microbial Community Genomics at the HOT/ALOHA; PBSM: “Pacific Beach Sand Metagenome” (Metagenoma da areia das praias do Pacífico); EBPR: “Phosphorus Removing Sludge Community” (Comunidade microbiana de plantas de estações de tratamento nos EUA e Austrália); DeepMed: “Mediterranean Bathypelagic Habitat Metagenome” (Metagenoma da comunidade microbiana da zona batipelágica do Mediterrâneo).



## 5.9 Análise filogenética das sequências ambientais homólogas às serina $\beta$ -lactamases

### 5.9.1 Classe A

#### 5.9.1.1 Esgotos hospitalares do Rio de Janeiro (EHRJ)

Duas árvores filogenéticas foram construídas com as sequências de prováveis  $\beta$ -lactamases classe A, conforme descrito no item 4.12. A primeira árvore contém as sequências dos esgotos hospitalares enquanto a segunda é formada pelas sequências prováveis de  $\beta$ -lactamases mineradas das bases CAMERA e IMG/M.

Na primeira árvore (Figura 5.11) podem ser observados três cladogramas principais. O clado 1 (com os ramos em vermelho) é formado por sequências das amostras hospitalares ZS e ZN, agrupadas próximas a sequências de  $\beta$ -lactamases classe A de *Enterobacter cloacae* (subclasse GES-1), *E. coli* (GES-5), *Geobacter* (subclasse y56), *Enterococcus* sp. (subclasse TEM) e *Klebsiella pneumoniae* (subclasse KPC) com a maioria dos valores de “bootstrap” acima de 70. O clado 2 (com ramos em azul) é o maior, compondo a maior parte da árvore, no qual as sequências presentes possuem uma distância filogenética mais curta entre elas (ramos curtos). Esse clado é formado por 92 sequências dos esgotos hospitalares, tanto das amostras ZS quanto de ZN, e se agruparam próximas de classe A de gêneros do filo Bacteroidetes como *Prevotella*, *Capnocytophaga* e *Bacteroides* e subclasses CfxA, CepA e CblA. O valor de “bootstrap” dos ramos mais internos do primeiro subclado foi baixo (< 50), porém o valor do ramo mais externo foi 98, indicando que pelo menos essas sequências hospitalares estão próximas, em um grande clado. Já nos subclados seguintes os valores de “bootstrap” foram altos (> 90). No terceiro clado (com ramos em roxo) sequências das réplicas 01 e 02 de ZN se agruparam próximas a uma classe A (subclasse VEB) de *Akkermansia muciniphila*, do filo Verrucomicrobia (Figura 5.11). As sequências dos esgotos hospitalares ZS e ZN estão com códigos alfa-numéricos ao lado, gerados pelo programa do pirosequenciador 454/Life Science.



### 5.9.1.2 Bases CAMERA e IMG/M

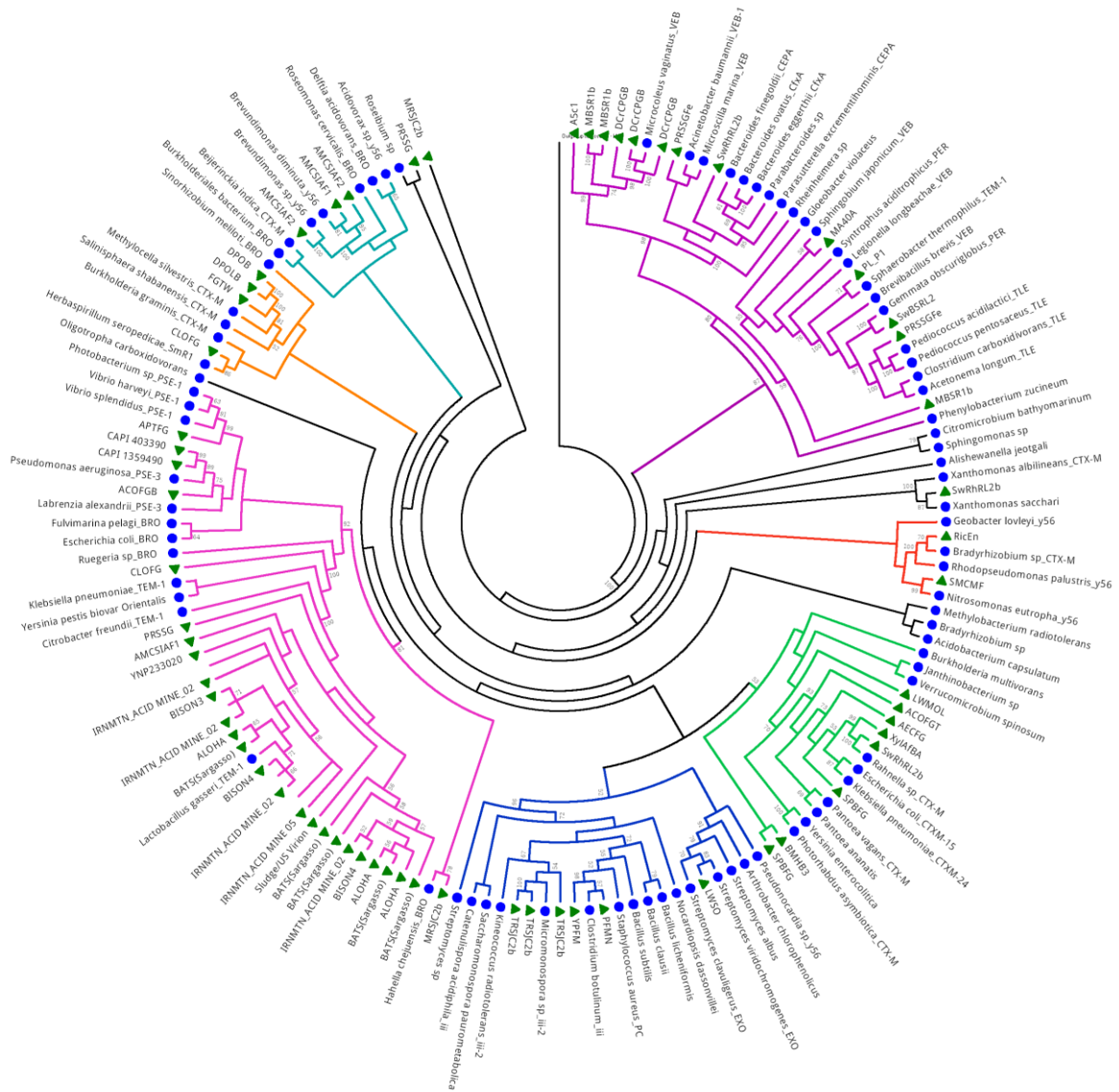
Na segunda árvore referente à classe A (Figura 12) podem ser observados sete clados principais, sendo que a maioria das sequências se agrupou mais próximas de sequências dos filos Proteobacteria e Firmicutes.

No primeiro grande clado (clado 1, ramos na cor amarela) foi formado por três subclados de sequências de diferentes projetos de ambientes terrestres (MBSR1b, DcrCPGB, A5c1,) sendo que as sequências que compõem o primeiro subclado agruparam próximas de uma  $\beta$ -lactamase de *Microcoleus vaginatus* (Cianobacteria). O segundo subclado também foi formado por sequências de projetos de ambientes terrestres (SwRhRL2b e PRSSGFe) e agruparam mais próximos de  $\beta$ -lactamases produzidas por bactérias pertencentes aos grupos Proteobacteria e Bacteroidetes. Já o terceiro subclado apresentou sequências de ambientes projetados (MA40A) e terrestres (PL\_P1, SwRhRL2b e PRSSGFe) que se agruparam próximas de  $\beta$ -lactamases produzidas por bactérias de diferentes filos como Proteobacteria, Chloroflexi, Planctomycetes e Firmicutes, respectivamente.

O clado 2 (ramos em vermelho) foi formado por sequências da base IMG/M, de projetos distintos de solo (RicEn e SMCMF). Cada uma se agrupou próxima de sequências produzidas por bactérias do filo Proteobacteria como *Bradyrhizobium sp.* e *Nitrosomonas eutropha*. Estas  $\beta$ -lactamases são pertencentes aos subgrupos CTX-M e y56. No clado 4 (ramos em salmão) apareceram sequências de quatro diferentes projetos metagenômicos de ambientes de solo como de comunidade microbiana de sedimento de um lago nos EUA (LWSO), projeto de palha de arroz (TRSJC2b), PFMN (microbioma fecal de suínos). Essas sequências se agruparam próximas a subclados distintos com sequências de  $\beta$ -lactamases classe A de actinobactérias como *Micromonospora sp.* de subclasse iii2 e *Streptomyces viridochromogenes* (subclasse EXO), e de *Clostridium botulinum* (subclasse iii). Já o clado 5 (ramos na cor rosa) foi composto por três subclados formados por prováveis sequências de  $\beta$ -lactamases classe A de projetos de metagenoma de ecossistemas bem distintos como ambientes terrestres (MRSJC2b e PRSSG), ambientes extremos (IRNMTN - “Acid Mine”, BISON3 e BISON4 - “Bison Metagenome”, e águas termais YNP233020), aquáticos como “ALOHA” e BATS (Mar de Sargasso, estação de Bermudas), de esgoto e de estação de tratamento de esgoto (“Sludge/US Virion”, CAPI e PLMO), de Ar (AMCSIAF1) e

interação da comunidade microbiana com artrópodes (ACOFGB, CLOFG e APTFG). No primeiro subclado a sequência de ambiente terrestre (MRSJC2b) se agrupou próxima à  $\beta$ -lactamase produzida por *Hahella chejuensis* (subclasse BRO), pertencente ao grupo Gammaproteobacteria. O segundo subclado foi formado por sequências de ambientes extremos, aquáticos, Ar e terrestre que agruparam próximas a *Citrobacter* (subclasse TEM-1), e *Lactobacillus gasseri* (subclasse TEM-1). No terceiro subclado as prováveis  $\beta$ -lactamases dos projetos metagenômicos se agruparam próximas de gêneros do grupo Gammaproteobacteria também, porém da subclasse PSE.

No clado 6 (ramos na cor laranja) observou-se que sequências de diferentes ambientes como da comunidade microbiana simbióticas com artrópodes como formigas (CLOFG), cupins (FGTW), e besouros (DPOB e DPOLB) que se agruparam próximas a sequências de  $\beta$ -lactamases classe A das subclasses CTX-M e BRO, produzidas por bactérias pertencentes ao filo Proteobacteria (Figura 12). O clado 7 (ramos em roxo) foi formado apenas por sequências de ambiente Ar, de Singapura (AMCSIAF1 e AMCSIAF2), que agruparam próximas de  $\beta$ -lactamases produzidas pelo gênero *Brevundimonas*, ambas da subclasse y56. Como havia muitas sequências com alta similaridade (acima de 97%) dos projetos “Acid Mine (IRMNT)” e “Bison Metagenome (BISON)” estas foram retiradas da árvore para deixá-la mais clara visualmente. A árvore foi calculada com 28 sequências do projeto “Acid Mine” e 14 do projeto “Bison Metagenome” (dos locais 1, 2, 3 e 4).



**Figura 5.12.** Filogenia das prováveis  $\beta$ -lactamases da classe A, recuperadas das bases de dados de projetos de metagenoma (CAMERA e IMG/M) que estão representadas com um triângulo fechado verde. O triângulo fechado azul representa as sequências de referência do grupo externo (classe D) colapsadas. A árvore filogenética foi calculada utilizando-se o programa RAXML-HPC, com o método de Máxima Verossimilhança e modelo de substituição WAG. Foram realizados “bootstraps” com 100 réplicas e seus valores podem ser observados nos ramos (acima de 50%). A barra de escala indica 0,5 substituições de aminoácidos por sítio. Os números de acesso das sequências representadas na árvore filogenética, assim como os das sequências metagenômicas, se encontram na sessão de Material Suplementar (em CD).

### 5.9.2. Classe C

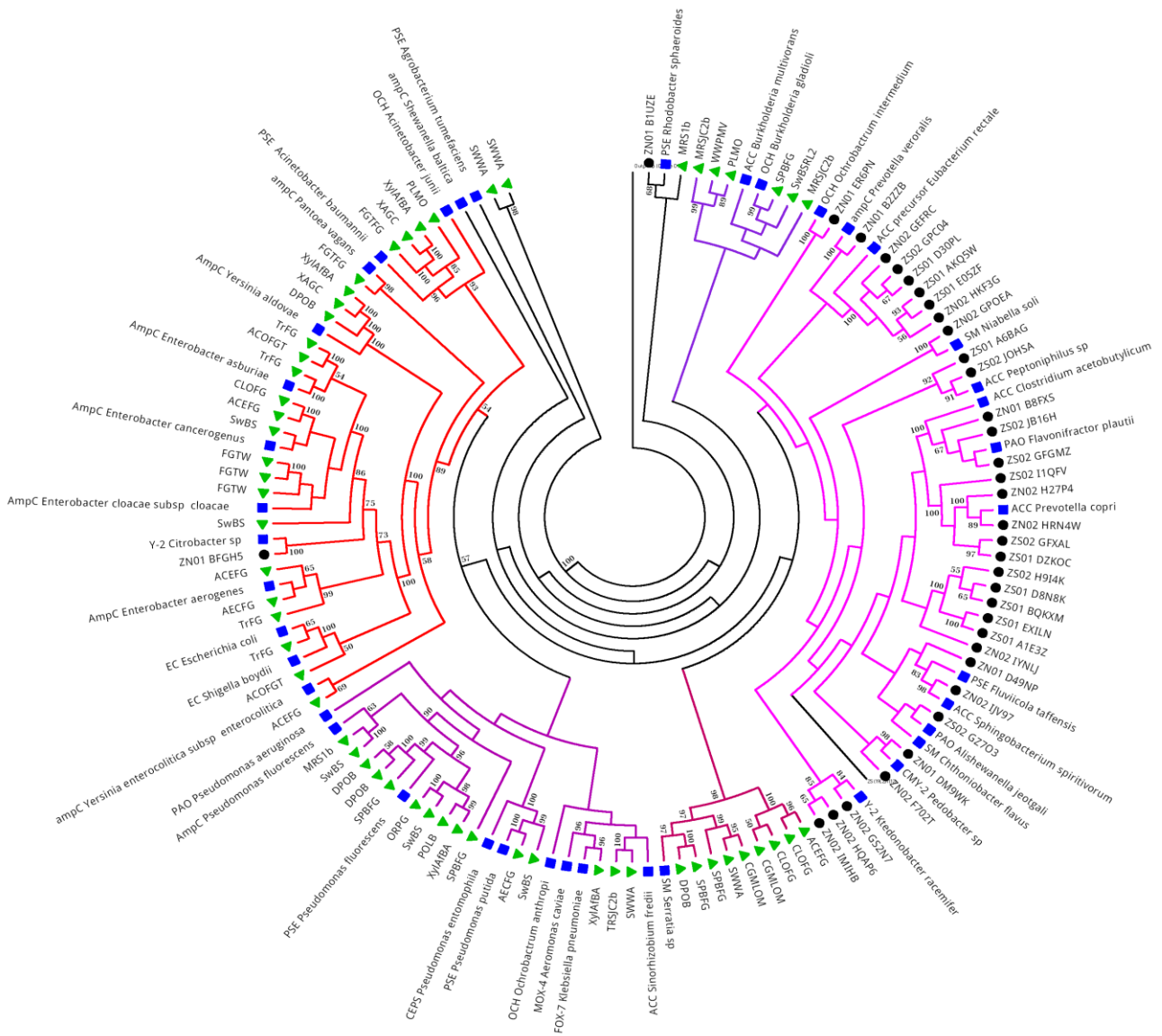
A filogenia das prováveis  $\beta$ -lactamases dos projetos metagenômicos do CAMERA e IMG/M agruparam com sequências da classe C, principalmente com as ampC (especialmente a subclasse ACC). Os ramos curtos entre sequências do mesmo clado sugerem uma distância evolutiva pequena e que possivelmente essas enzimas não divergiram muito entre si. As  $\beta$ -lactamases classe C representativas, sendo a maioria de origem não clínica (~87%), não ficaram muito distantes das  $\beta$ -lactamases encontradas em diferentes habitats de diferentes locais como: microbioma de fungos com interação com artrópodes como *Xyleborus affinis* (XylAfBA, XAGC, cupim), *Sirex noctillio* (SWWA, vespa), *Cyphomyrmex longiscapus* (CLOFG, formiga), *Dendroctonus frontalis* (SPBFG, besouro), projetos de estação de tratamento de esgoto na Suíça (PLMO, WWPMV), entre outros. Os projetos de microbiota de inseto, geralmente associados com bactérias simbióticas, foram a maioria, representando 74% (46 sequências) das prováveis  $\beta$ -lactamases de todos os projetos de metagenoma presentes na árvore. Apenas quatro sequências da classe C eram de projetos de tratamento de esgoto (6,4%) e as restantes eram de projetos relacionados a solo (19,3%) (Figura 5.13). Cinco clados principais podem ser visualizados, sendo o clado 1 (ramos em vermelho) formado por sequências de diferentes projetos de metagenoma de diferentes ambientes, que agruparam próximas a sequências de  $\beta$ -lactamases classe C subclasses ACC e OCH, de *Burkholderia multivorans* e *B. gladioli*, respectivamente.

O clado 2 (ramos na cor rosa forte), o maior clado, foi formado apenas por subclados de sequências dos esgotos hospitalares ZS e ZN. Essas sequências se agruparam próximas de  $\beta$ -lactamases classe C de diferentes subclasses como ACC, PAO, PSE, OCH, SM, CMY-2, Y-2 e ampC produzidas por bactérias pertencentes aos grupos Bacteroidetes e Firmicutes, principalmente, mas também com os grupos Proteobacteria (*Ochrobactrum intermedium*) e Chloroflexi (*Ktedonobacter racemifer*).

Os clados 3 (ramos em verde) e 4 (em roxo) foram formados apenas por sequências de projetos de metagenômica, de locais distintos, sendo a maioria de ambiente de interação fungo simbiótico-artrópode. No clado 3 os projetos de interação da comunidade microbiana com artrópodes (ACEFG, CLOFG, SPBFG, SWWA, DPOB e APTFG) se agruparam mais próximas de uma  $\beta$ -lactamase classe C subclasse SM, produzida por *Serratia* sp. (Proteobacteria).

Já no clado 4 (ramos em roxo) além dos projetos de interação de comunidade microbiana-artrópode, sendo a maioria besouros de diferentes gêneros (SWWA-vespa da madeira, XylAfBA, SPBFG, POLB e DPOB, besouro), além de sequências de solo (SwBS e TRSJC2b) e águas subterrâneas (ORPG). Estas sequências se agruparam próximas de  $\beta$ -lactamase classe C das subclasses ACC, OCH, PSE, CEPS, AmpC, PAO, MOX-4 (ESBL) e FOX-7, de bactérias pertencentes, em sua maioria, ao grupo Proteobacteria.

O clado 5 (ramos em salmão) foi formado basicamente por sequências de projetos metagenômicos, principalmente os relacionados à interação simbiótica da comunidade microbiana com artrópodes (ACEFG, ACOFGT, CLOFG e TrFG - formiga, FGTW e FGTFG - cupim, DPOB, XAGC e XylAfBA-besouro), mas também de ambientes de solo (SwBS) e de estação de tratamento de esgoto (PLMO). Essas sequências se agruparam próximas a  $\beta$ -lactamases classe C de subclasses distintas como OCH, AmpC, PSE, Y-2 e EC, porém a maioria sendo AmpC produzidas por bactérias pertencentes ao grupo Proteobacteria (Figura 5.13).



**Figura 5.13. Filogenia das prováveis  $\beta$ -lactamases da classe C, recuperadas das bases de dados metagenômicos CAMERA e IMG/M, e de esgotos hospitalares de amostras coletadas nos hospitais públicos da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro. As sequências representativas de  $\beta$ -lactamases obtidas no Refseq estão indicadas com um quadrado fechado azul ao lado, as sequências de projetos metagenômicos estão indicadas com um triângulo fechado verde e as sequências das amostras ZS e ZN estão indicadas com um círculo fechado preto. A árvore filogenética foi calculada utilizando o programa RAXML-HPC, com o método de Máxima Verossimilhança e modelo de substituição WAG. Foram realizados “bootstraps” com 100 réplicas e seus valores podem ser observados nos ramos (acima de 50%). A barra de escala indica 0,5 substituições de aminoácidos por sítio. Os números de acesso das sequências representadas na árvore filogenética encontram-se na sessão de Material Suplementar (em CD).**



### 5.9.2 Classe D

A árvore filogenética das prováveis sequências de  $\beta$ -lactamases classe D apresentou uma diversidade maior do que a da classe C, em relação aos projetos presentes nas bases CAMERA e IMG/M. Foram observados cinco clados principais, sendo o clado 1 (ramos em roxo) formado por sequências de diferentes projetos metagenômicos, englobando diferentes ambientes como os aquáticos, incluindo marinho (OSBM58 e Sargasso) e de água doce (LVIA5G), de estação de tratamento de esgoto (PLMO), de solo (SwBSRL2, PRSSGFe, MRS1b, PRSSG2), de ambientes extremos (BMHB, BMHB3), interação fungo-artrópode (ACODB e XAGC) e também da microbiota intestinal de humanos (HDG, HCG e SRS054956), além de sequências de amostras hospitalares ZN (01 e 02) e ZS (01 e 02). A maioria das sequências de esgotos hospitalares agrupou junta, em um subclado separado das sequências dos projetos de metagenoma, mais próximas de  $\beta$ -lactamases classe D OXA-10 de *Eubacterium rectale* e *Roseburia hominis*, bactérias comuns da microbiota intestinal humana.

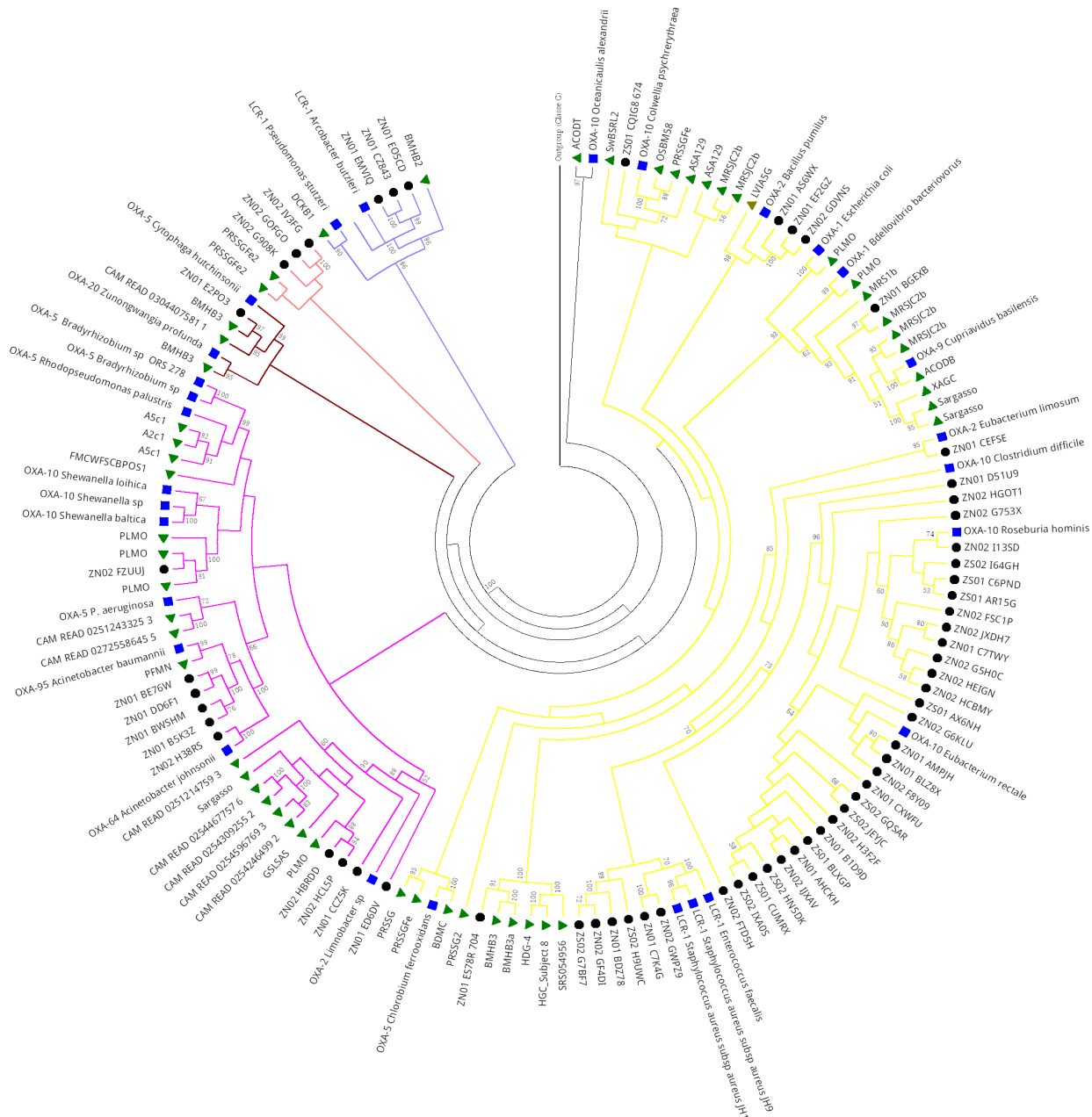
As sequências que compõem esse clado, no geral, agruparam próximas de  $\beta$ -lactamases classe D das variantes OXA-1, OXA-2, OXA-10, OXA-5 e LCR-1, sendo que as sequências de esgotos hospitalares ficaram mais próximas de OXA-10 e LCR-1, enquanto a maioria das sequências de projetos metagenômicos se agruparam mais próximas das restantes (OXA-1, OXA-2, OXA-5 e OXA-9) (Figura 5.14).

O clado principal 2 (ramos em rosa) também foi formado por sequências de diferentes ambientes de projetos de metagenoma como aquáticos (Sargasso, CAM\_READ, GSLSAS, FMCWFSCBPOS1), de estação de tratamento de esgoto (PMLO), de solo (A5c1, A2c1) e das amostras hospitalares de ZN01 e ZN02 (10 sequências no total). Essas sequências, em geral, se agruparam próximas de  $\beta$ -lactamases classe D OXA-10, OXA-64, OXA-95, e OXA-5.

O clado 3 (ramos em amarelo) é menor, formado por sequências de projeto de ambiente extremo (BMHB3), aquático marinho (CAM\_READ) e de esgotos hospitalares (ZN01) que se agruparam próximas de OXA-20 e OXA-5 de *Zunongwangia profunda* e *Cytophaga hutchinsonii*, respectivamente. Ambas as bactérias são do grupo Bacteroidetes.

O clado 4 (ramos em cor salmão) é pequeno, composto por sequências do efluente ZN02 e sequências do projeto de metagenoma relacionado a solo (PRSSGFe).

O clado 5 (ramos na cor azul claro) é composto por sequências de dois projetos metagenômicos de diferentes ambientes (DCKB e BMHB2-ambiente extremo), assim como por sequências de esgotos hospitalares da Zona Norte (ZN01). Essas sequências se agruparam próximas da subclasse LCR-1 de bactérias do grupo Proteobacteria (*P. stutzeri* e *Arcobacter butzleri*) (Figura 5.14). Vale ressaltar que as sequências dos esgotos hospitalares (ZS e ZN) agruparam mais próximas de sequências de  $\beta$ -lactamases classe D representativas do filo Firmicutes, enquanto as sequências dos projetos de metagenoma agruparam mais próximas de  $\beta$ -lactamases classe D representativas do filo Proteobacteria.



**Figura 5.14. Filogenia das prováveis sequências de  $\beta$ -lactamases da classe D, recuperadas das bases de dados metagenômicos CAMERA, IMG/M e dos esgotos hospitalares de amostras coletadas nos hospitais públicos da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro. As sequências representativas de  $\beta$ -lactamases estão indicadas por um quadrado fechado azul, as sequências dos projetos de metagenoma estão indicadas por um triângulo fechado verde e as sequências dos esgotos hospitalares do Rio estão indicadas por um círculo fechado preto. O grupo externo (classe C) está indicado por um triângulo preto que representa sequências colapsadas. A árvore filogenética foi calculada com o programa RAXML-HPC, com o método de Máxima Verossimilhança e modelo de substituição WAG. Foram realizados “bootstraps” com 100 réplicas e seus valores podem ser observados nos ramos (acima de 50%). A barra de escala indica 0,5 substituições de aminoácidos por sítio. Os números de acesso das sequências representadas na árvore filogenética encontram-se na sessão de Material Suplementar (em CD).**

## 5.10 Caracterização do metagenoma de esgotos hospitalares do Rio de Janeiro

### 5.10.1 Análise dos dados gerados a partir da amplificação do 16S ribossomal bacteriano

#### 5.10.1.1 Curva de rarefação

A curva de rarefação obtida a partir da amplificação de parte do gene 16S ribossomal (rrs) das amostras de esgotos hospitalares ZS e ZN indicou que não foi possível esgotar a diversidade das amostras em relação à espécie. Esse resultado foi preliminar para estimar a diversidade das amostras e avaliar a quantidade de corridas necessárias para o pirosequenciamento realizado posteriormente. A partir da curva mostrada na Figura 5.15 percebe-se que 250 sequências não foram suficientes para a curva atingir o platô, indicando uma grande diversidade nas amostras, sobretudo na amostra ZN.

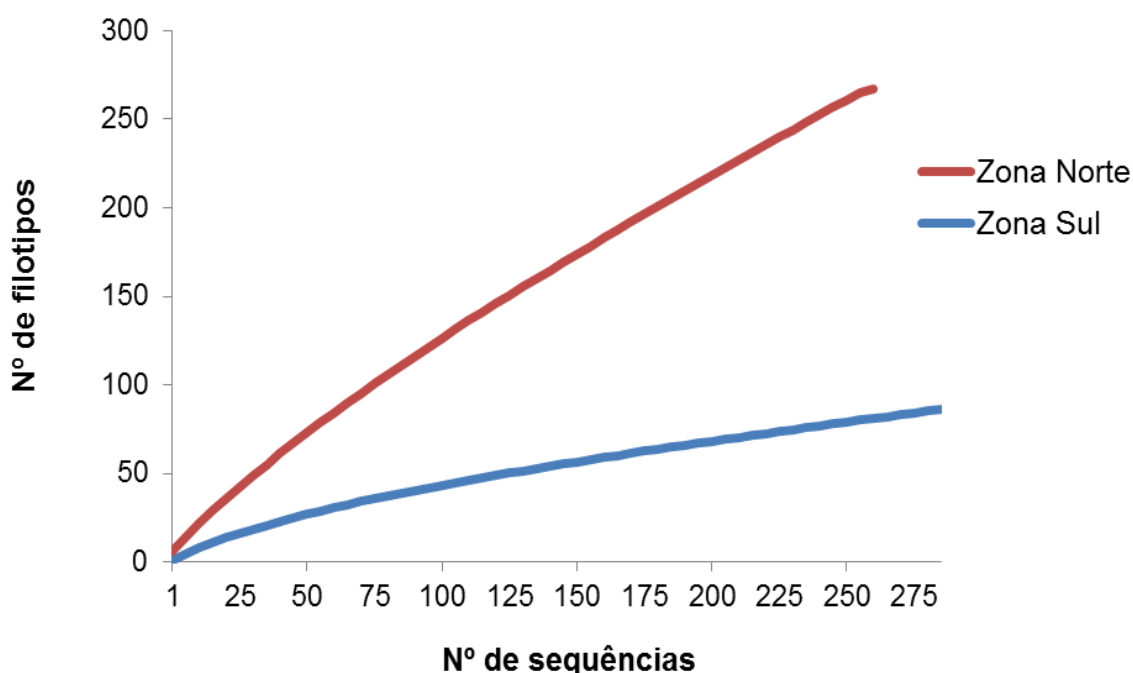
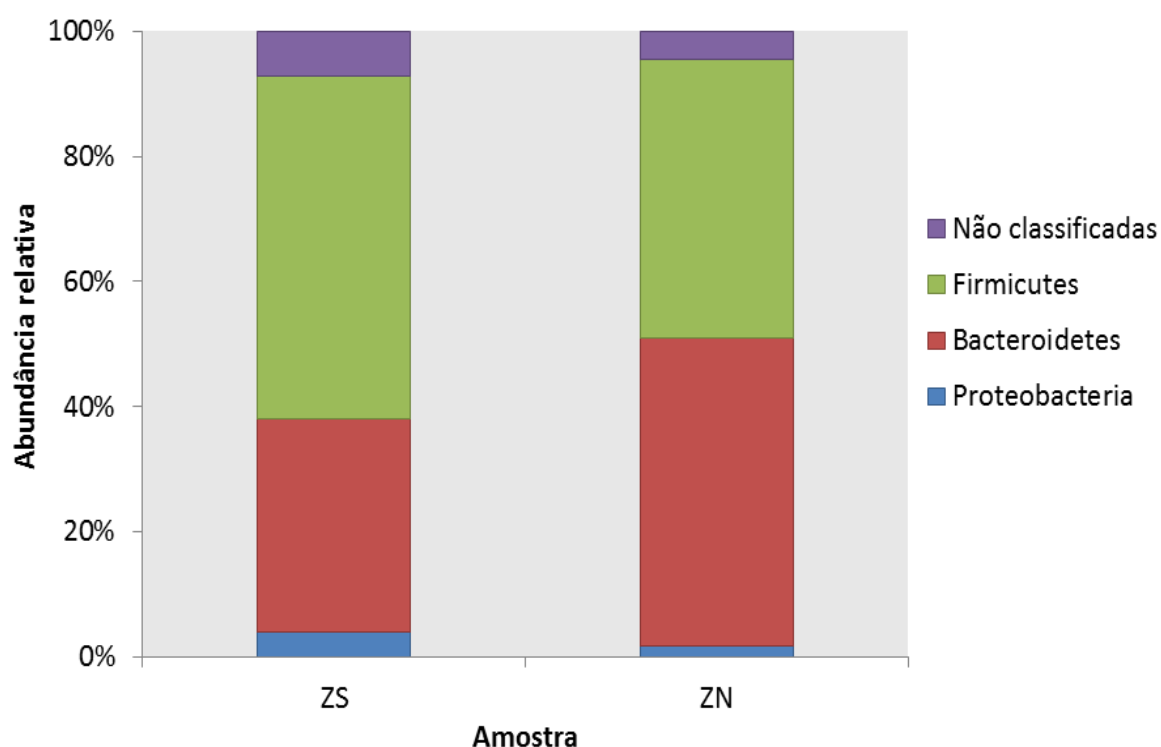


Figura 5.15. Curva de rarefação. Sequências parciais do gene 16S ribossomal (rrs) de Bacteria foram utilizadas para estimar a diversidade bacteriana nas amostras de esgotos hospitalares do Rio, ZS e ZN, baseado na identidade de sequências de 97% como ponto de corte para o delineamento de unidades taxonômicas operacionais (UTOs) pelo MOTHUR. ZS: Hospital Zona Sul. ZN: Hospital Zona Norte.

### 5.10.1.2 Diversidade taxonômica a partir do gene 16S ribossomal

A partir da amplificação por PCR de parte do 16S ribossomal (rrs) dos esgotos hospitalares ZS e ZN, pode-se observar que a abundância relativa dos filos predominantes não variou muito entre os dois hospitais. Em ambas as amostras hospitalares predominaram Firmicutes (55% e 44% para ZS e ZN, respectivamente) e Bacteroidetes (34% e 49% para ZS e ZN, respectivamente). Já Proteobacteria representou uma pequena parcela das sequências classificadas, com aproximadamente 7% e 4,5% para ZS e ZN, respectivamente (Figura 5.16).

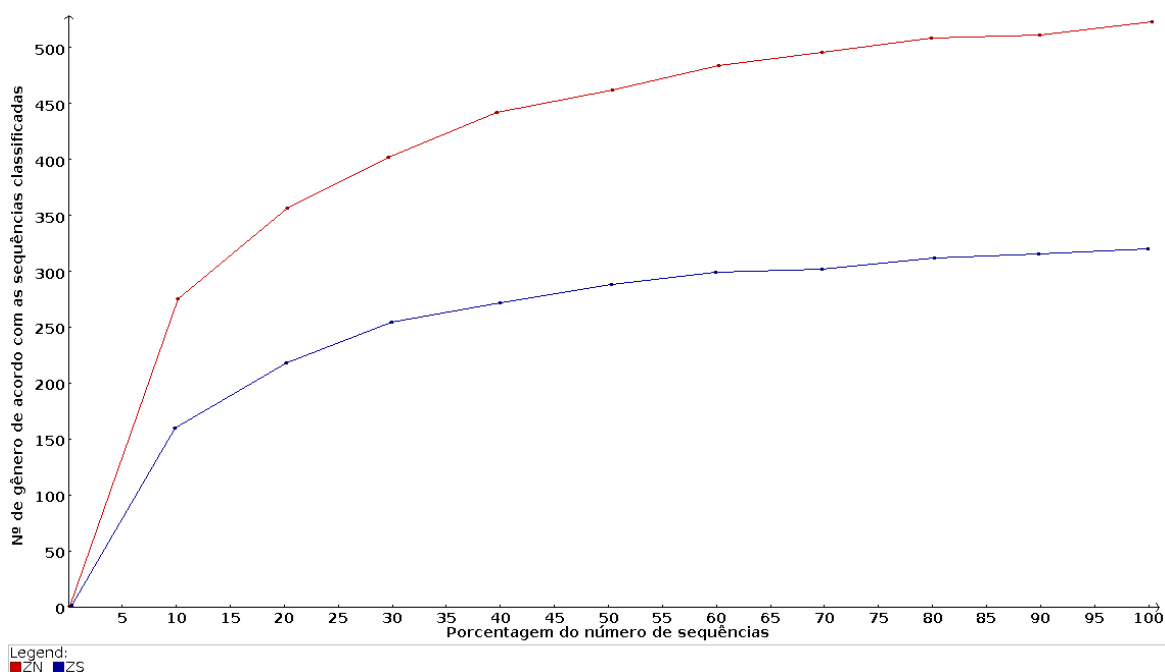


**Figura 5.16** Abundância relativa (%) dos filos mais predominantes nas amostras de esgotos hospitalares da Zona Sul (ZS) e Zona Norte (ZN) do Rio de Janeiro a partir da amplificação do gene ribossomal 16S.

## 5.10.2 Dados metagenômicos dos EHRJ obtidos a partir do pirosequenciamento

### 5.10.2.1 Curva de rarefação

A curva de rarefação gerada pelo MEGAN, ao nível de gênero revelou que a diversidade em Bacteria foi esgotada para ambas as amostras ZS e ZN. Porém, a diversidade da amostra ZS se aproximou mais do platô com um número menor de sequências quando comparado com a curva de ZN. Enquanto a curva da amostra ZS estabilizou em aproximadamente 300 tipos de gêneros de Bacteria, a curva de ZN estabilizou pouco acima dos 500 gêneros (Figura 5.17).

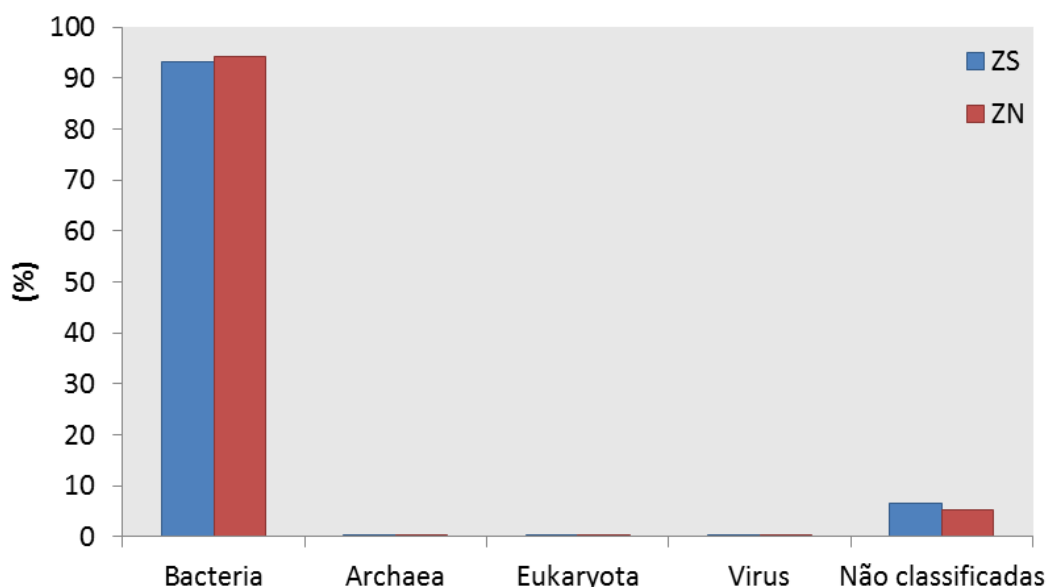


**Figura 5.17** Curva de rarefação das amostras ZN e ZS, localizados no Rio de Janeiro-RJ, gerada a partir do MEGAN, de acordo com as sequências pertencentes ao grupo Bacteria que foram classificadas ao nível de gênero, a partir dos resultados do BLASTN. ZS: Hospital localizado na Zona Sul. ZN: Hospital localizado na Zona Norte.

### 5.10.2.2 Visão geral da distribuição da comunidade microbiana

Como uma visão geral da composição microbiana, no momento da coleta, presente nos esgotos dos dois hospitais públicos do Rio de Janeiro pode-se observar que domínio Bacteria foi predominante, com 93% para ZS e 94% para ZN, como esperado em função da metodologia utilizada. Eukarya, Vírus e Archea apareceram em proporções baixas: 0,02% na amostra ZS e

0,06% na amostra ZN para Archea; 0,13% em ZS e 0,22% em ZN para Eukaryota; e 0,12% em ZS e 0,15% em ZN para vírus. Quanto às sequências não caracterizadas foram 7% para ZS e 5,2% para ZN, respectivamente (Figura 5.18). Como os resultados das réplicas 01 e 02 de cada hospital foram similares então foi descrito um único valor para ambas.



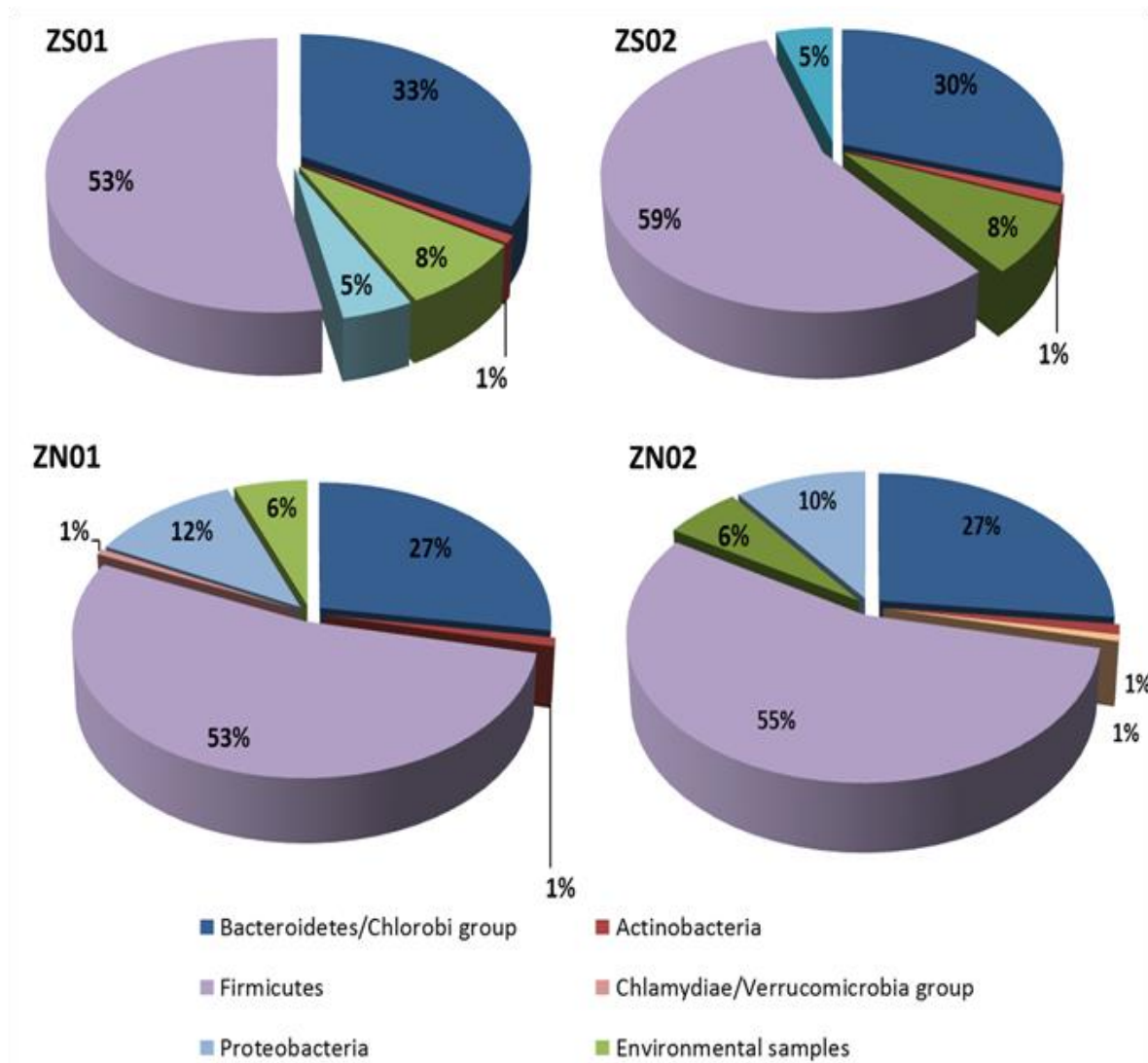
**Figura 5.18** Diversidade taxonômica das amostras de esgotos hospitalares ZS e ZN localizados no Rio de Janeiro-RJ. ZS: Hospital da Zona Sul. ZN: Hospital da Zona Norte. Os resultados das réplicas 01 e 02 de cada hospital estão representados juntos.

### 5.10.2.3 *Diversidade taxonômica da comunidade bacteriana*

De forma a aprimorar o conhecimento do conteúdo genético da comunidade microbiana dos esgotos hospitalares foi utilizada uma abordagem de metagenômica comparativa a partir de dados de pirosequenciamento. O resultado da distribuição das sequências revelou uma predominância dos filos Firmicutes (53 e 56% para as réplicas ZS01 e ZS02 e 53 e 55% para as réplicas ZN01 e ZN02, respectivamente), Bacteroidetes (33 e 30% para ZS01 e ZS02, respectivamente e 27% para ambas as réplicas de ZN) e Proteobacteria (5% para ZS01 e ZS02 e 10 e 12% para ZN01 e ZN02, respectivamente).

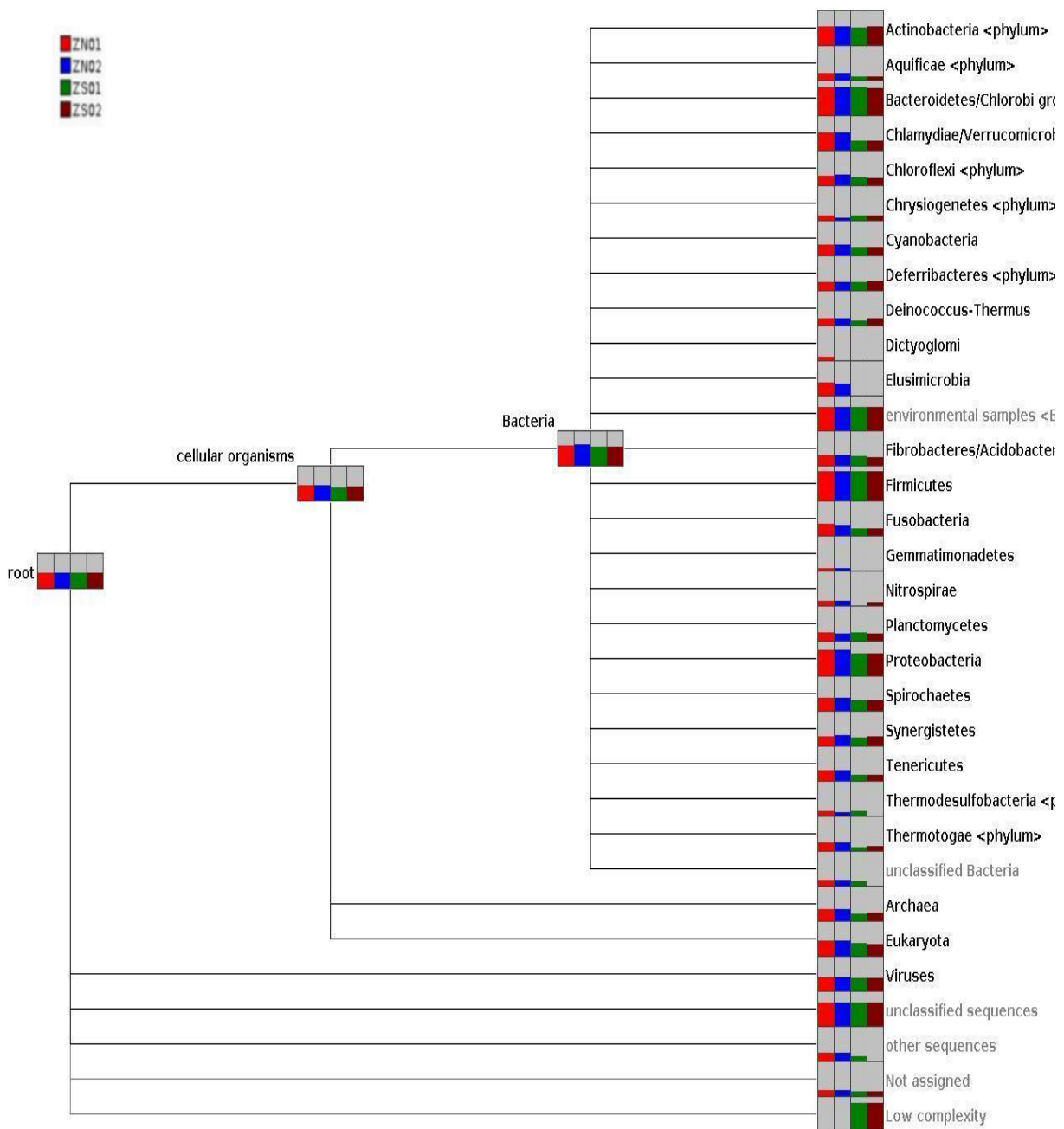
Dentre esses, Firmicutes foi o grupo taxonômico mais abundante de todos (Figuras 5.19 e 5.20), envolvendo mais da metade das sequências. De qualquer forma, o perfil da comunidade bacteriana das duas amostras de esgoto hospitalar, em relação ao filo, classe, ordem e família apresentaram porcentagens de abundância relativas muito semelhantes. Além disso, o

número de bactérias não classificadas também foi bem similar para as duas amostras, com 38 e 42% para ZS e ZN, respectivamente.



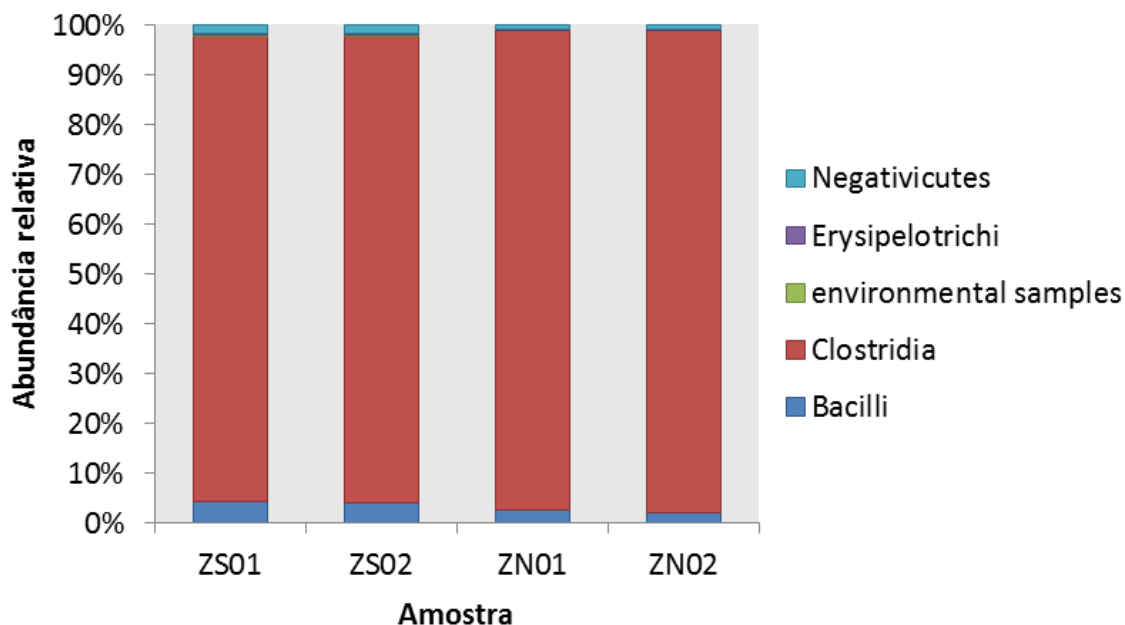
**Figura 5.19.** Gráfico da distribuição taxonômica de Bacteria presente nas amostras dos esgotos hospitalares da Zona Sul (ZS01, ZS02) e da Zona Norte (ZN01 e ZN02). Nos gráficos acima estão representados apenas os grupos taxonômicos predominantes.





**Figura 5.20. Gráfico em barras de cada réplica das amostras hospitalares gerado pelo programa MEGAN 4. ZN01 (vermelho): Hospital da Zona Norte réplica 01. ZN02 (azul): Hospital da Zona Norte réplica 02. ZS01 (verde): Hospital da Zona Sul réplica 01. ZS02 (vinho): Hospital da Zona Sul réplica 02. Esses dados foram normalizados. A altura das barras corresponde ao número de “hits” para cada táxon. “Not assigned” indica sequências que tiveram uma correspondência com a base de dados de nucleotídeos NT (NCBI) que não foram atribuídas a um táxon. “No hits” indica as sequências que não apresentaram resultado do BLASTN significativo (“E-value” >10<sup>-5</sup>). O nodo de “Low complexity” representa pequenas sequências com baixa identidade.**

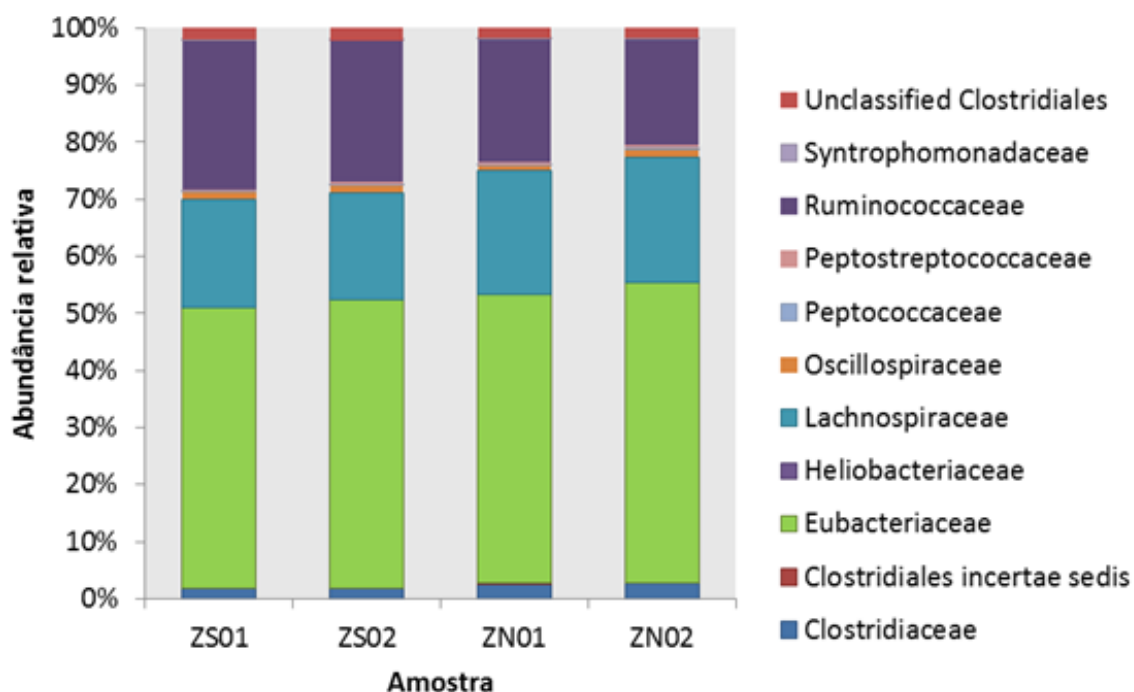
Considerando o filo Firmicutes, as bactérias mais abundantes encontradas foram Clostridia com 97 e 94% para ZN e ZS, respectivamente e em menor abundância Bacilli com aproximadamente 2,3 e 4% para ZN e ZS, respectivamente. Os grupos Negativicutes (1 e 1,7 % para ZN e ZS, respectivamente) e Erysipelotrichi 0,22 e 0,35% para ZN e ZS, respectivamente, foram grupos também presentes, mas em uma abundância muito baixa em comparação com o grupo predominante (Figura 5.21).



Ordem (Filo Firmicutes)	Frequência relativa (%)			
	ZS01	ZS02	ZN01	ZN02
<i>Bacilli</i>	4,2	4,00	2,57	2,01
<i>Clostridia</i>	93,37	93,65	96,3	96,7
<i>Negativicutes</i>	1,71	1,66	0,88	1,02
<i>Erysipelotrichi</i>	0,36	0,34	0,012	0,015
<i>Environmental samples</i>	0,36	0,34	0,22	0,22

**Figura 5.21** Distribuição taxonômica (gráfico) e frequência relativa (tabela) do filo Firmicutes presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro, RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

Em relação à classe Clostridia, 99% das sequências pertenciam à ordem Clostridiales, para ambas as amostras. Considerando Clostridiales os grupos mais abundantes para ZN e ZS, respectivamente, foram: Eubacteriaceae (52 e 50%), Ruminococcaceae (22 e 26%) e Lachnospiraceae (20 e 19%). Os grupos com baixa predominância para ZN e ZS, respectivamente, foram: Clostridiaceae (2,6 e 1,8%), Oscillospiraceae (1,1 e 1,4%), Peptostreptococcaceae (0,42 e 0,25%), Peptococcaceae (0,21 e 0,15%), Clostridiales incertae sedis (0,07 e 0,04%), Clostridiales não classificados (1,8 e 2,1%), Heliobacteriaceae (0,03 e 0,02%) e Syntrophomonadaceae (0,03 e 0,01%) (Figura 5.22).



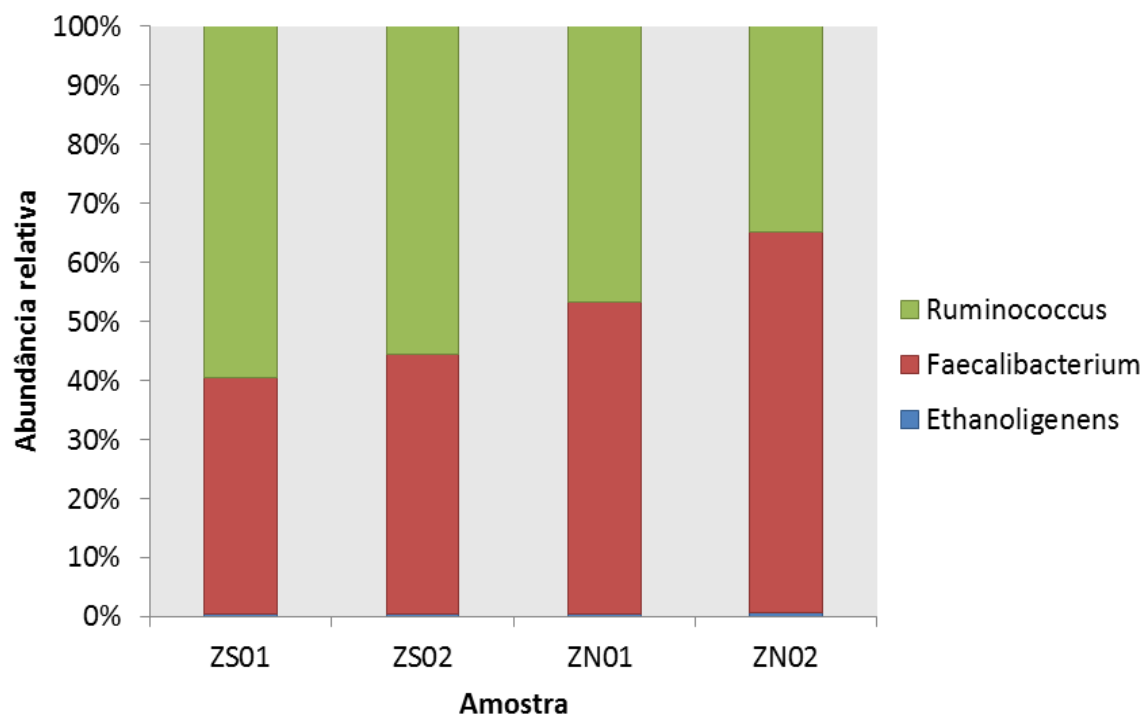
**Figura 5.22** Distribuição taxonômica do grupo Clostridiales presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro, RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

Na família Eubacteriaceae (Filo Bacteroidetes) o grupo mais abundante foi *Eubacterium*, representando 99,95% e 99,93% nas amostras ZS e ZN, respectivamente, do total de sequências que tiveram uma atribuição da classificação dentro deste grupo. Em menor abundância foi o *Acetobacterium*, com 0,047% e 0,067% em ZS e ZN, respectivamente (Tabela 5.10).

**Tabela 5.10. Frequência relativa das sequências pertencentes à família Eubacteriaceae, presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

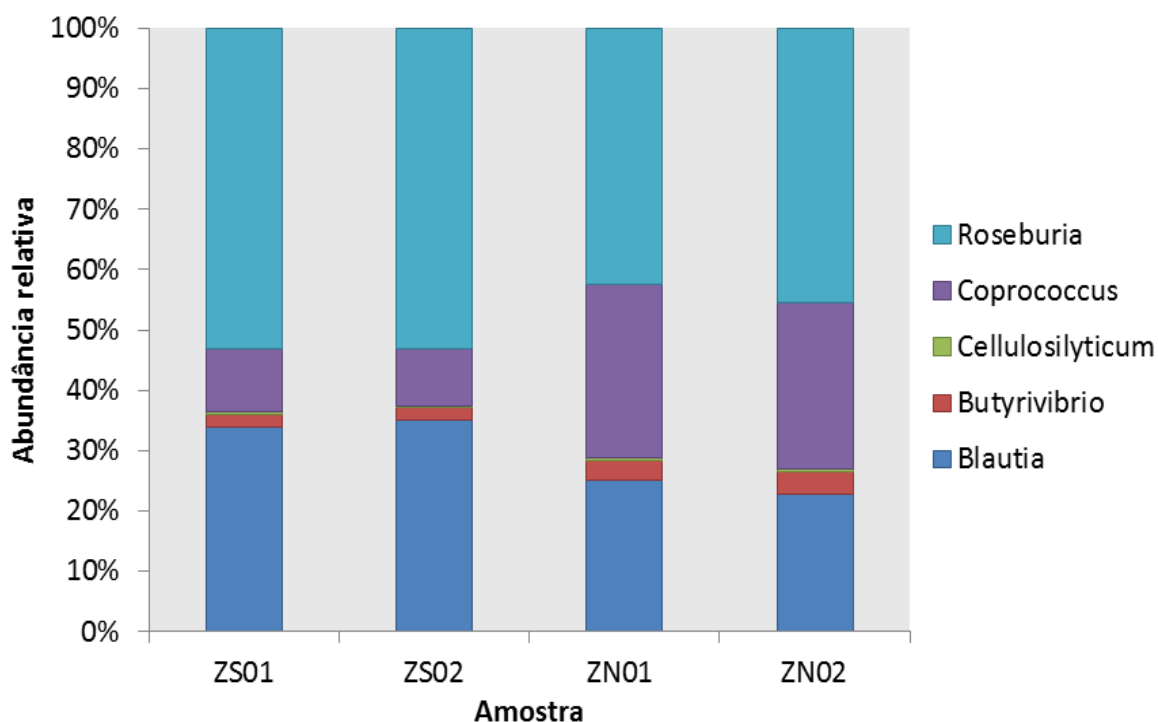
Gênero pertencentes à Eubacteriaceae	Frequência relativa (%)			
	ZS01	ZS02	ZN01	ZN02
<i>Eubacterium</i>	99,95	99,95	99,93	99,93
<i>Acetobacterium</i>	0,05	0,04	0,07	0,06

Em relação à família Ruminococcaceae os grupos mais abundantes foram *Ruminococcus* (60% e 40% para ZS e ZN, respectivamente) e *Faecalibacterium* (40% e 60% para ZS e ZN, respectivamente), e em menor abundância foram Ethanoligenens, com 0,46% e 0,6% em ZS e ZN, respectivamente (Figura 5.24).



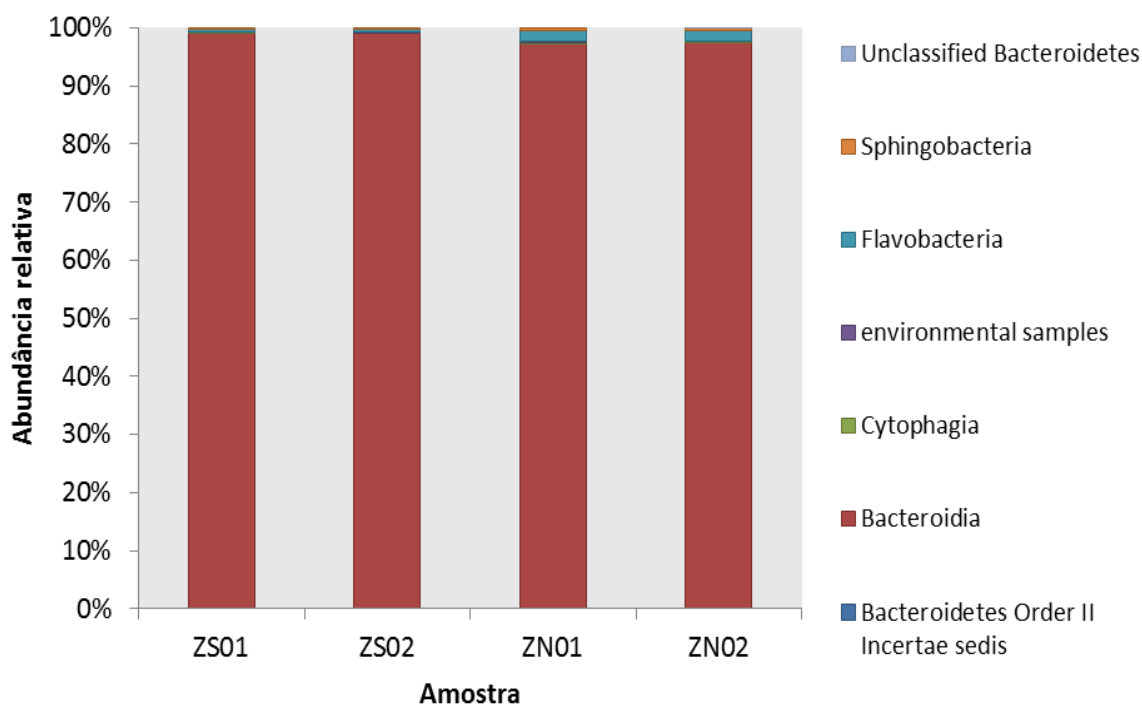
**Figura 5.23** Distribuição taxonômica das sequências pertencentes à família Ruminococcaceae, presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

Em relação à família Lachnospiraceae, os gêneros mais abundantes observados foram *Roseburia* (53% e 44% para ZS e ZN, respectivamente), *Blautia* (34% e 24% para ZS e ZN, respectivamente) e *Coprococcus* (10% e 28% para ZS e ZN, respectivamente), sendo as sequências relacionadas aos primeiros dois gêneros (*Roseburia* e *Blautia*) mais abundantes nas amostras ZS, enquanto as sequências relacionadas ao gênero *Coprococcus* foi mais abundante em ZN. Em menor abundância, foram também detectados os gêneros *Butyrivibrio* (2,1% e 3,5 para ZS e ZN) e *Cellulosilyticum* (0,3% e 0,4% para ZS e ZN) (Figura 5.25).



**Figura 5.24** Distribuição taxonômica das sequências pertencentes à família Lachnospiraceae, presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

Considerando o filo Bacteroidetes, a classe mais abundante foi Bacteroidia com 97% na amostra ZN e 99% na amostra ZS (para ambas as réplicas). Em menor abundância, foram as classes Sphingobacteria (0,34% e 0,5%), Flavobacteria (0,35 e 2%), Cytophagia (0,1 e 0,3%) para ZS e ZN, respectivamente (as duas réplicas apresentaram resultados semelhantes). Bacteroidetes Order II *Incertae sedis* foi encontrado, em pouca abundância (0,015%), apenas na réplica 02 da amostra ZS e em ambas as réplicas de ZN, na mesma proporção, porém também em pouca abundância (0,02%) (Figura 5.26).

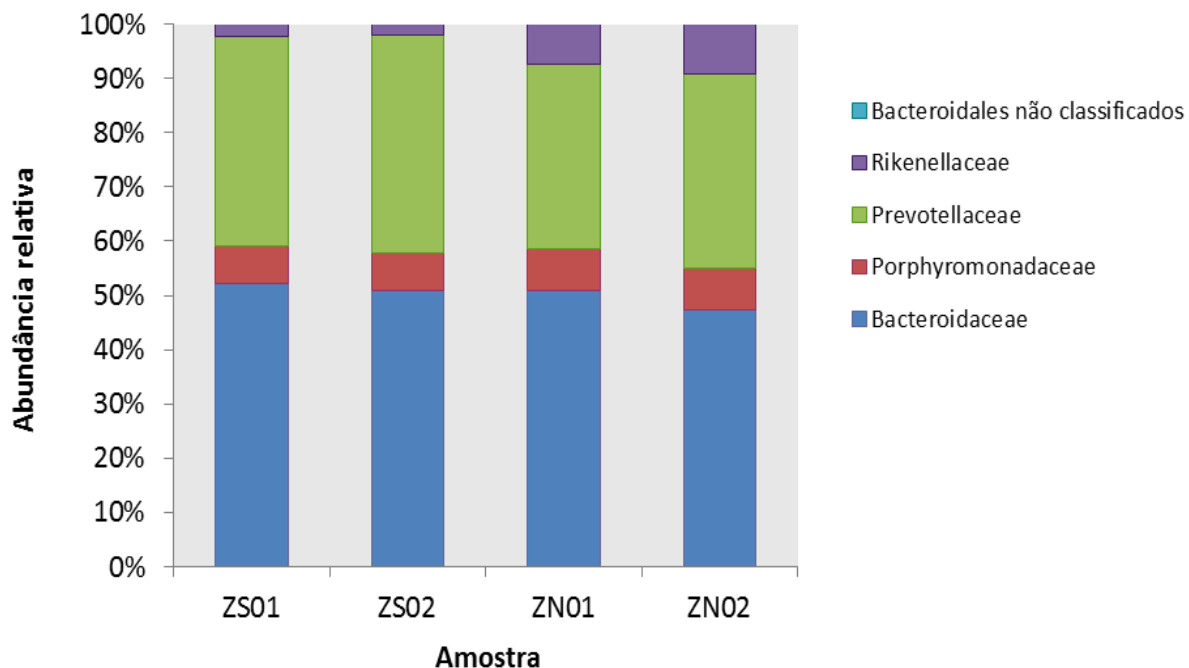


Classe (Filo Bacteroidetes)	Frequência relativa (%)			
	ZS01	ZS02	ZN01	ZN02
Unclassified Bacteroidetes	0,00	0,00	0,00	0,01
<i>Sphingobacteria</i>	0,34	0,32	0,49	0,51
<i>Flavobacteria</i>	0,35	0,40	1,88	1,72
<i>Environmental samples</i>	0,09	0,1	0,07	0,08
<i>Cytophagia</i>	0,11	0,15	0,27	0,28
<i>Bacteroidia</i>	99,08	99,03	97,27	97,39
<i>Bacteroidetes Order II Incertae sedis</i>	0,01	0,00	0,02	0,02

**Figura 5.25. Distribuição taxonômica (gráfico) e valores da frequência relativa (tabela) do filo Bacteroidetes presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

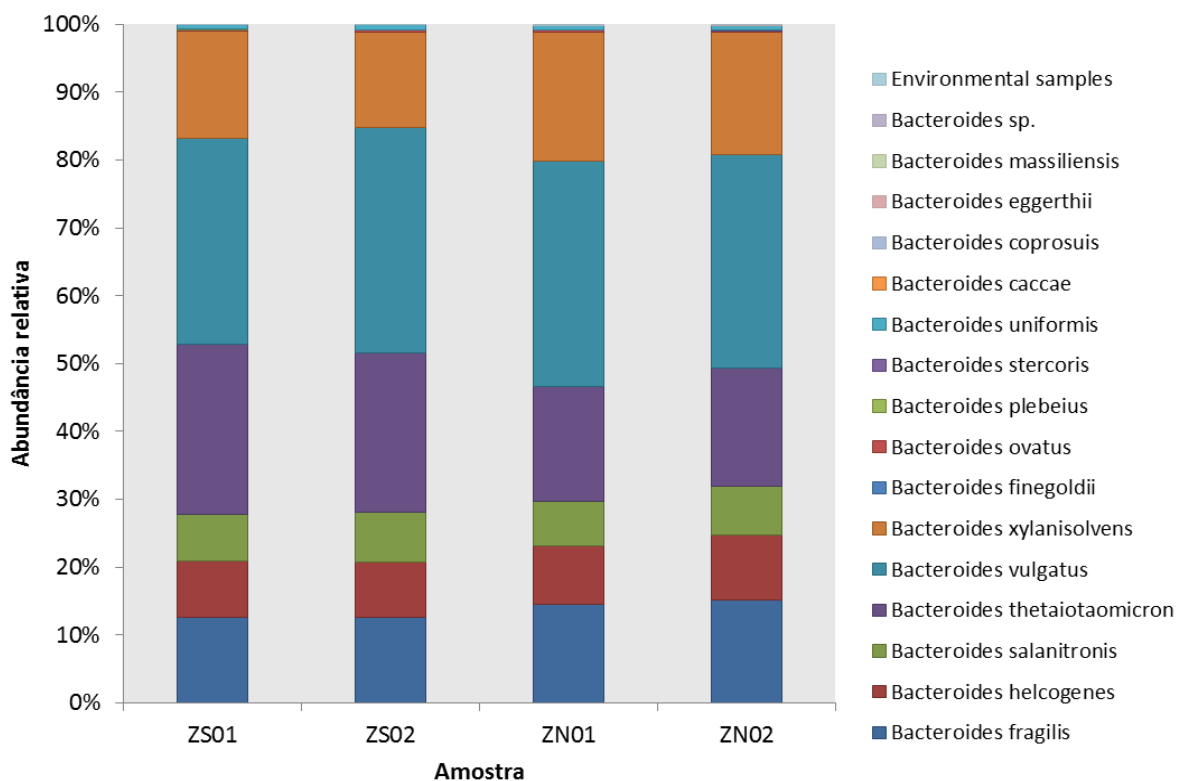


Considerando Bacteroidales, apenas apareceram sequências pertencentes ao grupo Bacteroidales e, nesse grupo, os mais abundantes para ZS e ZN, respectivamente, foram: Bacteroidaceae (51 e 49%) e Prevotellaceae (39 e 35%), e em menor abundância foram Porphyromonadaceae (7 e 8%) e Rikenellaceae (2 e 8%) (Figura 5.27).



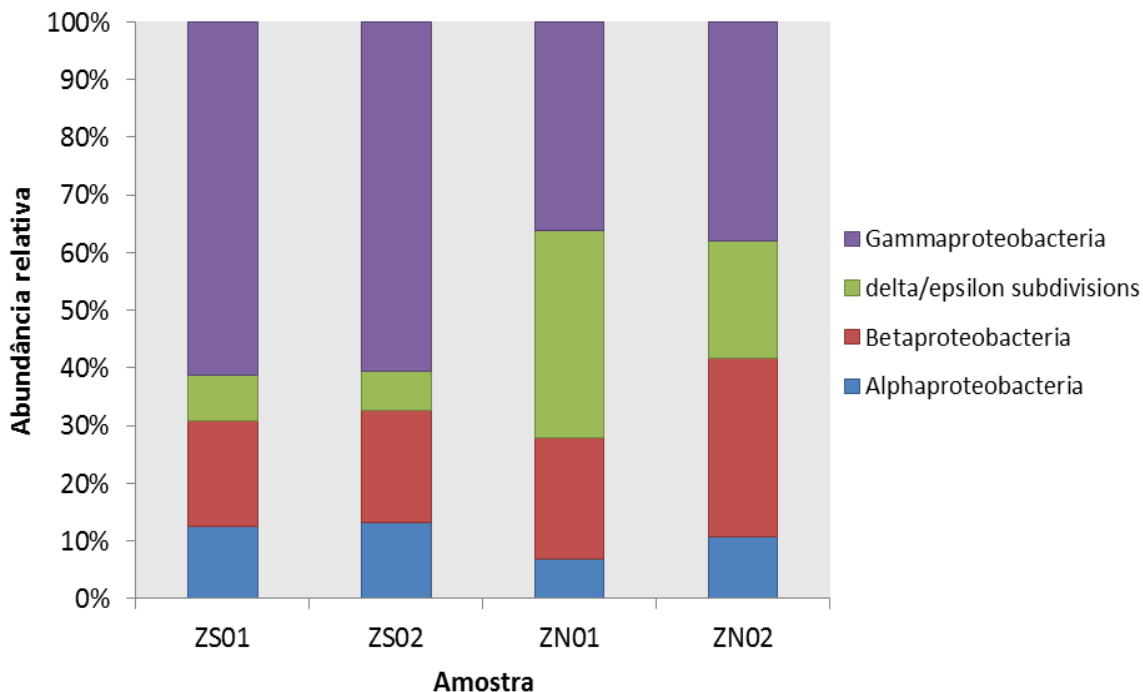
**Figura 5.26 Distribuição taxonômica da ordem Bacteroidales presente nas amostras de esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

Considerando a família Bacteroidaceae o gênero mais abundante foi Bacteroides, sendo as espécies *B. vulgatus*, *B. thetaiotaomicron* e *B. fragilis* as mais prevalentes em ambas as réplicas 01 e 02 do hospital da Zona Sul (ZS01 e ZS02) com aproximadamente 30, 25 e 13%. Já as réplicas do hospital da Zona Norte (ZN01 e ZN02) e as espécies *B. vulgatus*, *B. xylanisolvans* e *B. fragilis*, com valores bem próximos entre as réplicas, de aproximadamente 33, 17 e 15%, respectivamente (Figura 5.28).



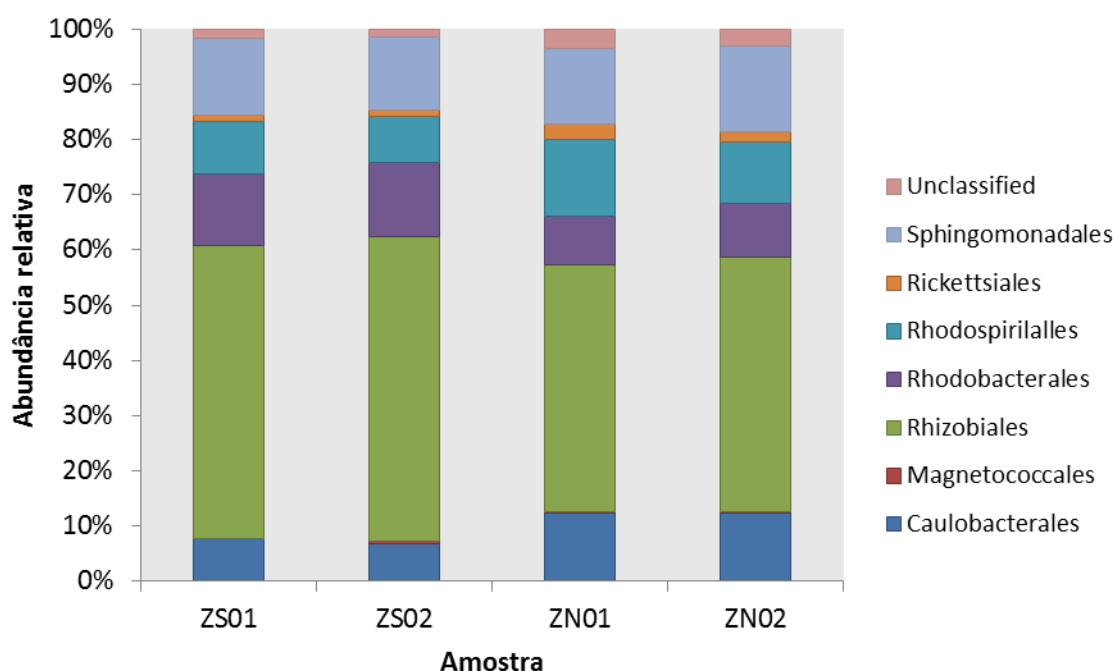
**Figura 5.27** Abundância relativa dos gêneros do filo Bacteroides presentes nas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

Em relação à Proteobacteria, as classes mais abundantes para as amostras de ZS e ZN, respectivamente foram: Gammaproteobacteria (60 e 36%), Betaproteobacteria (18 e 28%), Alphaproteobacteria (12 e 8%) e as subdivisões Delta/Epsilon-Proteobacteria (8 e 28%) (Figura 5.29).



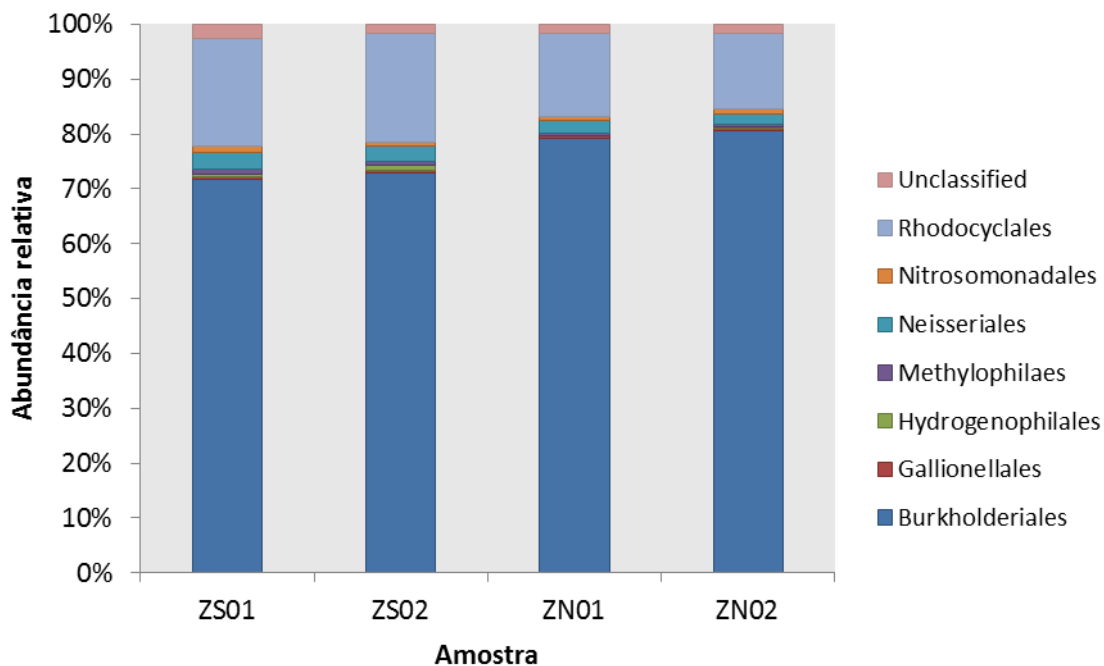
**Figura 5.28 Distribuição taxonômica do filo Proteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

Em relação à classe Alphaproteobacteria, o grupo mais abundante para ZN e ZS foi Rhizobiales com 46 e 54%, respectivamente. Em segundo lugar predominou Sphingomonadales com 15 e 14% e em terceiro houve uma diferença entre as amostras dos dois hospitais. A amostra ZN apresentou 12,2% para Caulobacterales e Rhodospirillales, enquanto a amostra ZS apresentou maior abundância para Rhodobacterales (14%) e em terceiro lugar Rhodospirillales com 9% (Figura 5.30). Como as réplicas 01 e 02 de cada hospital apresentaram uma distribuição muito similar entre si, apenas um valor de porcentagem foi mencionado.



**Figura 5.29** Distribuição taxonômica da classe Alphaproteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

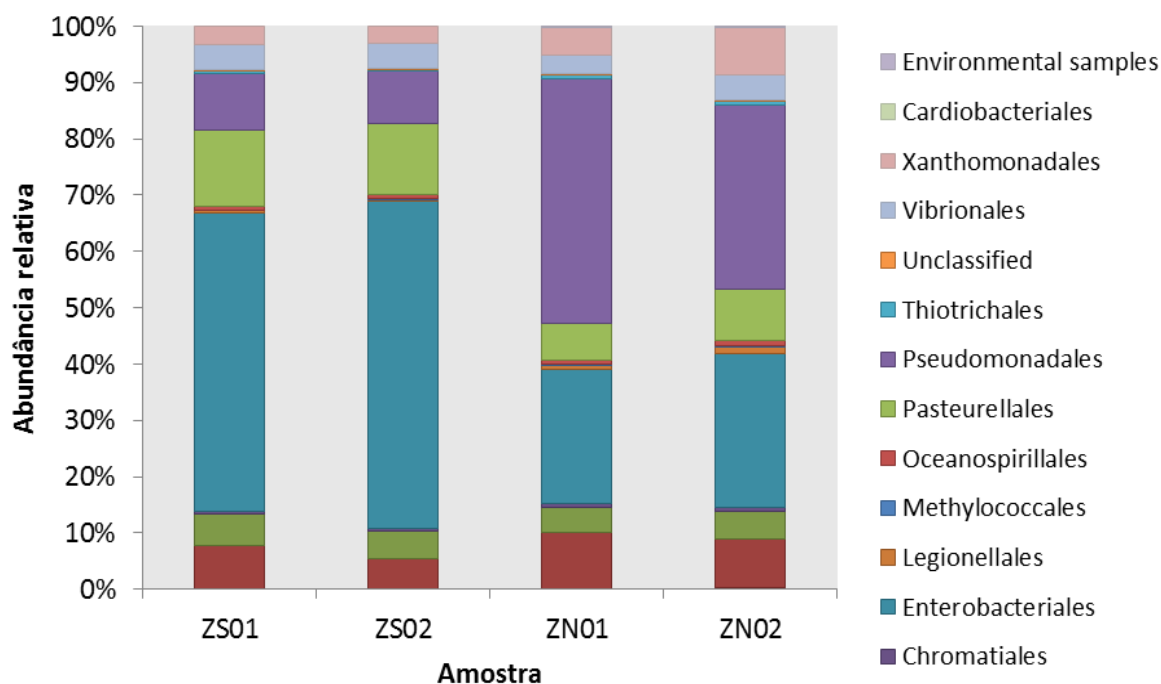
Em relação à classe Betaproteobacteria, os grupos mais abundantes para ZS e ZN foram Burkholderiales com 71 e 80%, respectivamente. O segundo grupo predominante foi Rhodocyclales com 19 e 14%, respectivamente. As demais classes (Nitrosomonadales, Neisseriales, Methylophilales, Hydrogenophilales e Gallionellales) apresentaram frequências muito baixas em ambos os hospitais (Figura 5.31).



Classe (Filo Proteobacteria)	Frequência relativa (%)			
	ZS01	ZS02	ZN01	ZN02
Não classificado	2,59	1,74	1,66	1,76
<i>Rhodocyclales</i>	19,57	19,69	15,11	13,69
<i>Nitrosomonadales</i>	1,19	0,73	0,71	0,81
<i>Neisseriales</i>	3,07	2,94	2,32	1,98
<i>Methylophilales</i>	0,91	0,60	0,47	0,43
<i>Hydrogenophilales</i>	0,42	0,87	0,16	0,23
<i>Gallionellales</i>	0,56	0,53	0,31	0,47
<i>Burkholderiales</i>	71,7	72,9	79,3	80,6

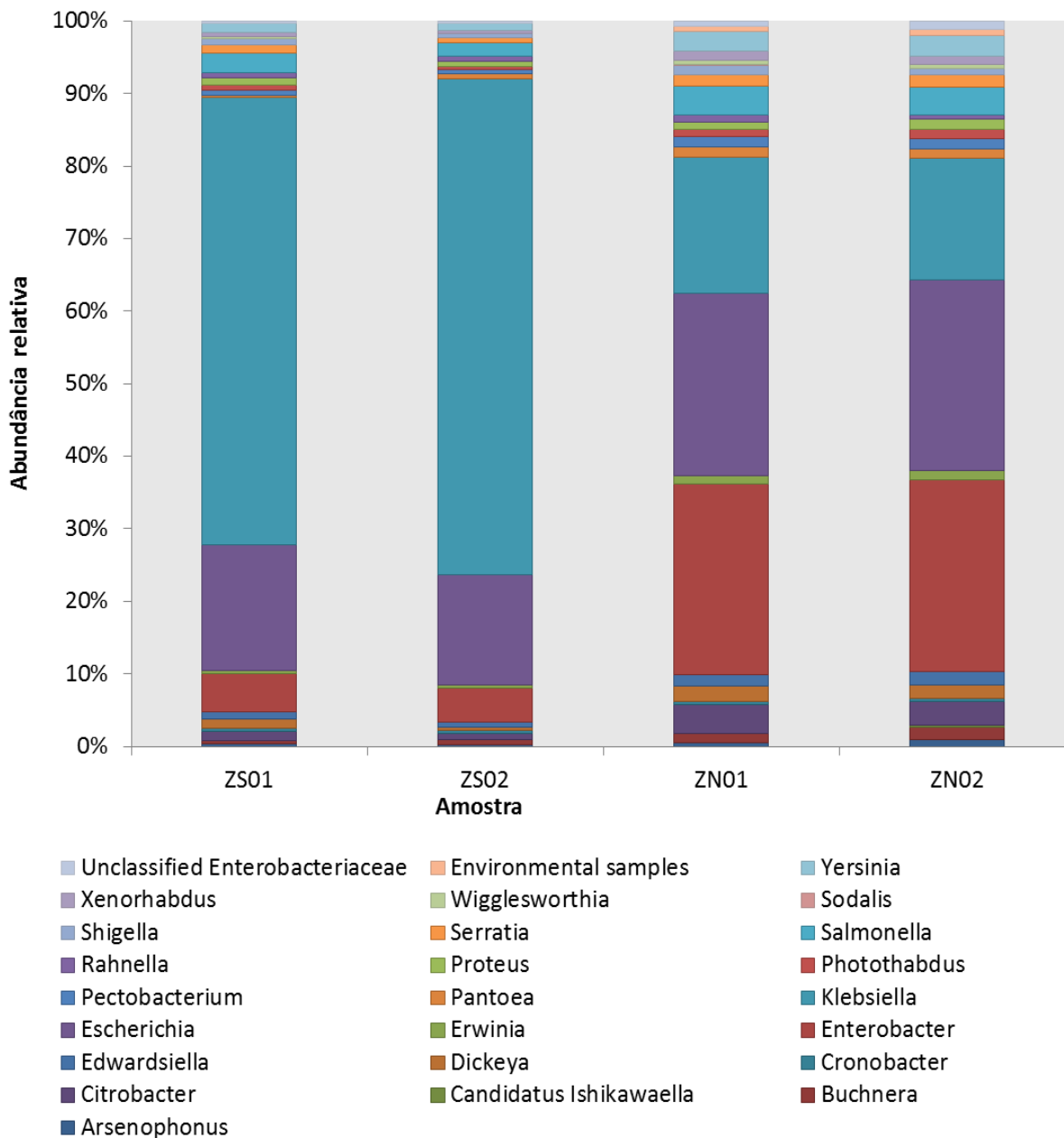
**Figura 5.30. Distribuição taxonômica (gráfico) e valor da frequência relativa (tabela) da classe Betaproteobacteria presente nas duas amostras dos esgotos hospitalares, localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

Em relação ao grupo Gammaproteobacteria, a família Enterobacteriales foi a mais abundante nas amostras ZS, com 56%, enquanto que Pasteurellales ficou em segundo lugar com 13%, seguido da família Pseudomonadales, com 9,7%. As amostras ZN apresentaram um resultado diferente, sendo a família mais abundante a Pseudomonadales, com 39%, seguido da Enterobacteriales com 25% e em terceiro Aeromonadales, com 9,4% (Figura 5.32).



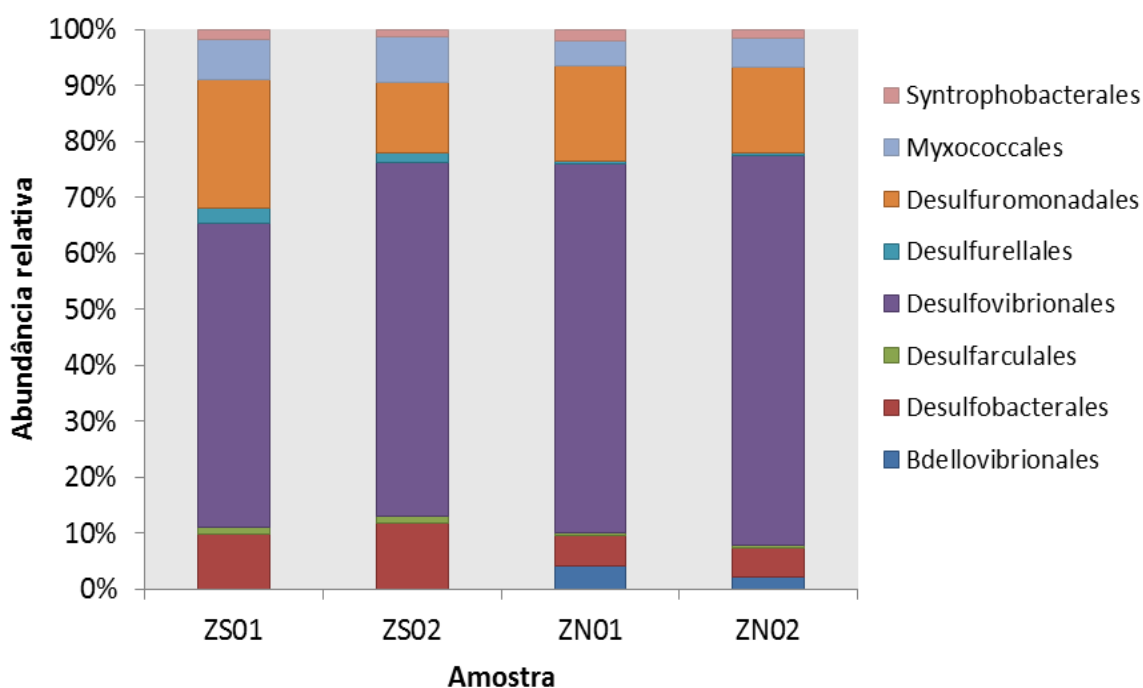
**Figura 5.31. Distribuição taxonômica da classe Gammaproteobacteria presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

Considerando a família Enterobacteriaceae, os gêneros mais abundantes nas amostras de esgotos hospitalares ZS e ZN foram: *Klebsiella*, *Escherichia* e *Enterobacter*. Nas amostras ZS01 e ZS02 a abundância de *Klebsiella* foi maior que nas amostras ZN, apresentando 61,6% e 68% em ZS01 e ZS02, respectivamente e 19% e 17% em ZN01 e ZN02, respectivamente. *Escherichia* e *Enterobacter* foram mais abundantes nas amostras ZN01 e ZN02, com 25% e 26% (*Escherichia*) e 26% e 27% (*Enterobacter*) em relação a ZS01 e ZS02, que apresentaram 17% e 15% de *Escherichia* e 5% de *Enterobacter*, respectivamente (Figura 5.33).



**Figura 5.32** Abundância relativa dos gêneros da família Enterobacteriaceae presentes nas amostras dos esgotos hospitalares localizados no Rio de Janeiro, RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.

No que concerne o grupo Delta-Proteobacteria também houve diferença dos grupos mais abundantes entre as amostras hospitalares. As amostras ZN01 e ZN02 apresentaram uma abundância maior de representantes da família Desulfovibrionales em relação às amostras ZS01 e ZS02, com 66 e 68%, respectivamente, enquanto em ZS01 e ZS02 houve uma abundância relativa de 54 e 63%, respectivamente. Considerando Desulfuromonadales a abundância foi aproximadamente homogênea entre as quatro amostras, com 23 e 13% para ZS01 e ZS02, respectivamente, e 17 e 15% para ZN01 e ZN02, respectivamente. Em relação às famílias Desulfobacterales e Myxococcales, as amostras ZS01 e ZS02 apresentaram uma maior abundância com 10 e 12% (Desulfobacterales) e 7% e 8,3% (Myxococcales), respectivamente, em relação a ZN01 e ZN02, que apresentaram 5%. A família Bdellovibrionales só apareceu nas amostras de esgotos do hospital da Zona Norte (ZN) com 4 e 3%, respectivamente (Figura 5.34).



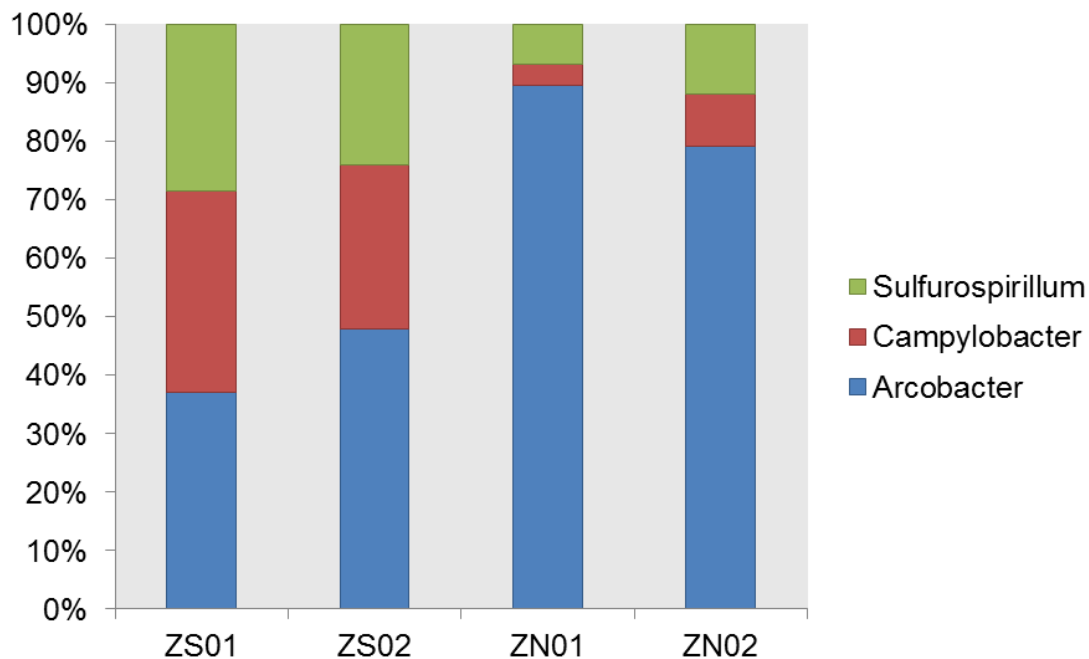
**Figura 5.33** Distribuição taxonômica da classe Delta-Proteobacteria presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.



No que concerne o grupo Epsilon-Proteobacteria, houve predominância de membros da Ordem Campylobacterales, com 100% nas amostras ZS01 e ZS02 e 99,4 e 99% em ZN01 e ZN02, respectivamente. A Ordem Nautiliales apareceu apenas nas amostras ZN, com 0,14% (ZN01) e 0,18% (ZN02) (Tabela 5.11). Em relação à Campylobacterales, o gênero mais abundante nas quatro amostras foi o *Arcobacter*, sendo mais prevalente nas amostras ZN01 (90%) e ZN02 (79%). Os gêneros *Campylobacter* e *Sulfospirillum* foram mais abundantes nas amostras do hospital da Zona Sul, com 34% e 28% (ZS01) e 29% e 24% (ZS02). Essas informações estão representadas graficamente na Figura 5.34.

**Tabela 5.11. Frequência relativa das Ordens pertencentes à classe Epsilonproteobacteria presente nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

Ordem (Classe Epsilonproteobacteria)	Frequência relativa (%)			
	ZS01	ZS02	ZN01	ZN02
Environmental samples	0,00	0,00	0,05	0,16
Unclassified	0,00	0,00	0,44	0,54
Epsilonbacteria				
Nautiliales	0,00	0,00	0,14	0,18
Campylobacterales	100	100	99,38	99,12



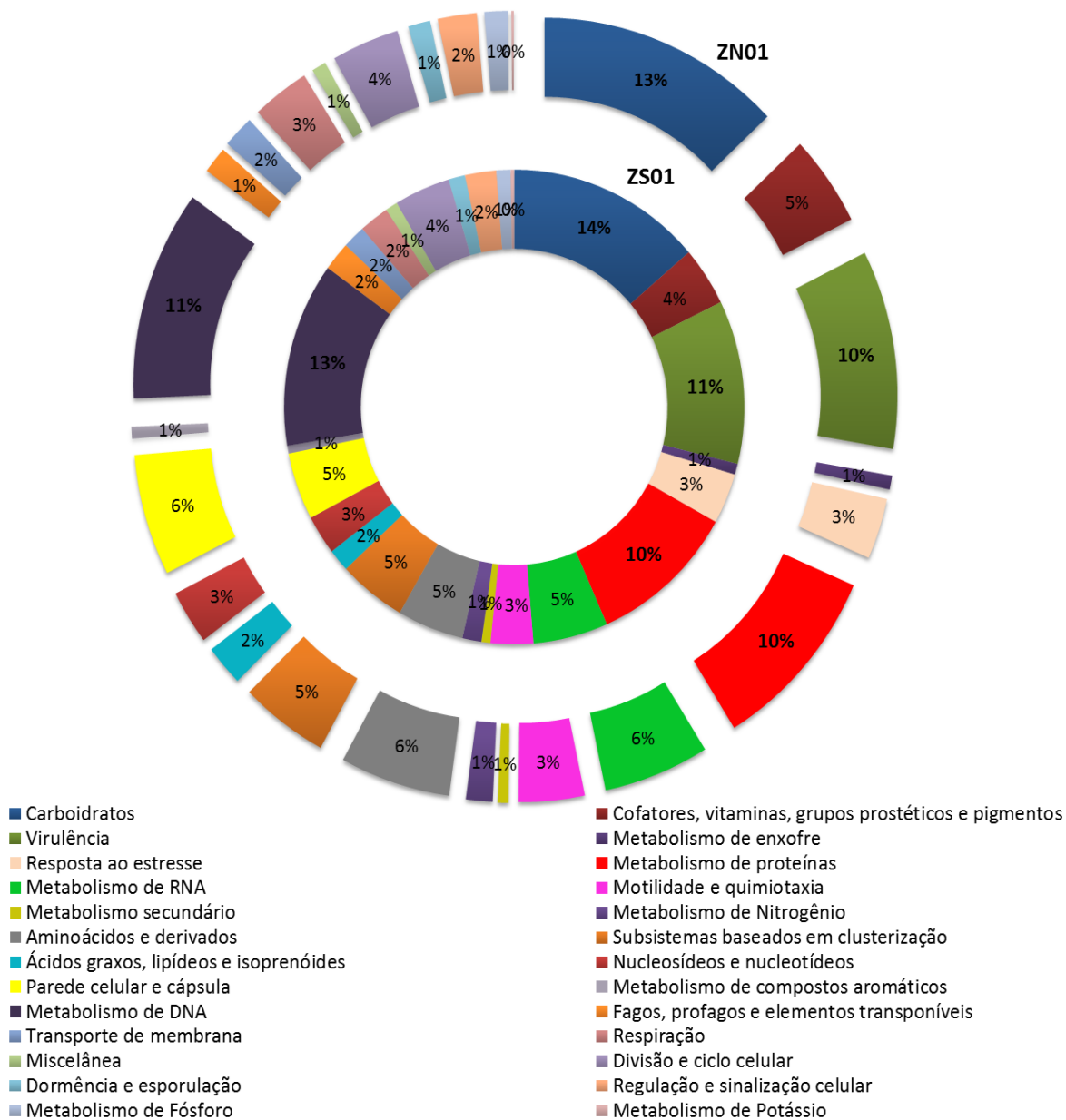
**Figura 5.34. Distribuição taxonômica dos gêneros pertencentes à Ordem Campylobacterales presentes nas amostras de esgotos hospitalares localizados no Rio de Janeiro-RJ. ZS01: Hospital da Zona Sul, réplica 01; ZS02: Hospital da Zona Sul, réplica 02. ZN01: Hospital da Zona Norte, réplica 01; ZN02: Hospital da Zona Norte, réplica 02.**

De uma maneira geral a distribuição taxonômica de ambas as amostras foram homogêneas em relação aos grandes grupos, principalmente entre as réplicas de cada hospital. Porém, algumas diferenças importantes foram observadas entre os hospitais em relação à abundância de Enterobacterales e Pseudomonadales e entre *Klebsiella* e *Enterobacter*.

### 5.11 Perfil do potencial metabólico do metagenoma

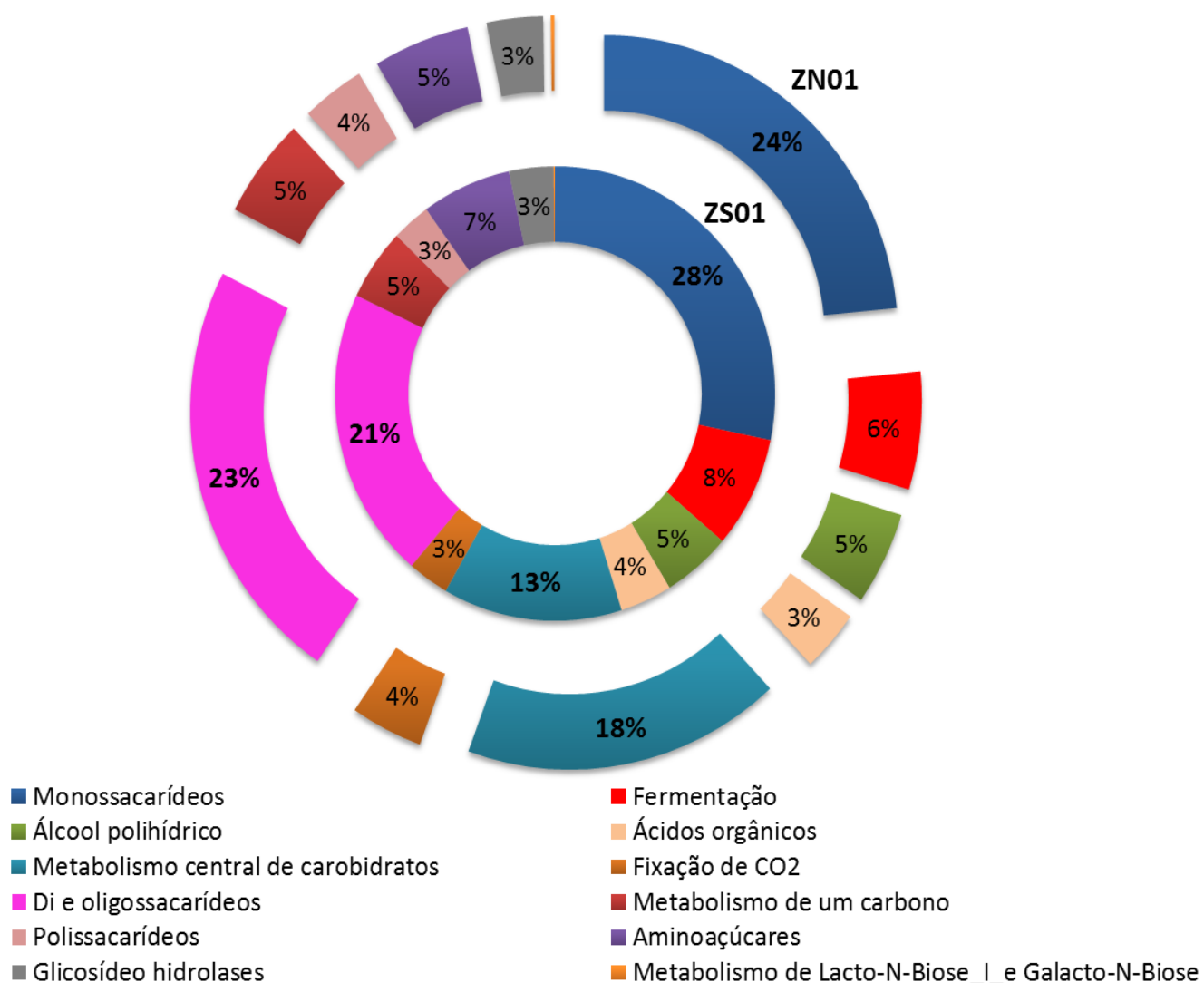
Os resultados do BLASTX apresentados são relacionados apenas para as réplicas 01 de cada hospital (ZS01 e ZN01). A partir da análise do BLASTX contra o Refseq de proteínas, 446.784 “hits” (93.7%) foram encontrados. Apenas 6,3% das sequências não apresentaram “hit” com nenhuma sequência da base a partir do valor de “E-value” selecionado ( $10^{-5}$ ). Destas sequências o programa MEGAN4 foi capaz de classificar apenas 25.774 das 326.574 sequências proteicas da amostra ZS01, de acordo com a base de dados SEED. Em relação às sequências que tiveram uma função atribuída (25.774), a maioria estava relacionada a subsistemas de carboidratos (13,4%), metabolismo de DNA (13%), virulência (11,2%) e metabolismo de

proteínas (11%) (Figura 5.35). Em relação à réplica 01 da amostra hospitalar da Zona Norte (ZN01), a análise do BLASTX contra a base Refseq resultou em 675.400 “hits” (96,6%) e apenas 3,4% das sequências não apresentaram nenhum “hit” de acordo com o valor de “E-value” selecionado ( $10^{-5}$ ). Destes 96,6%, o MEGAN4 foi capaz de classificar apenas 113.094 sequências proteicas da amostra ZN01, de acordo com a base de dados SEED e a maioria (562.306 sequências) não apresentou uma classificação funcional (“Not assigned”). As sequências que tiveram uma função atribuída (113.094) apresentaram, em maior abundância, os mesmos subsistemas observados para a amostra ZS01, com valores de porcentagens similares (Figura 5.35).



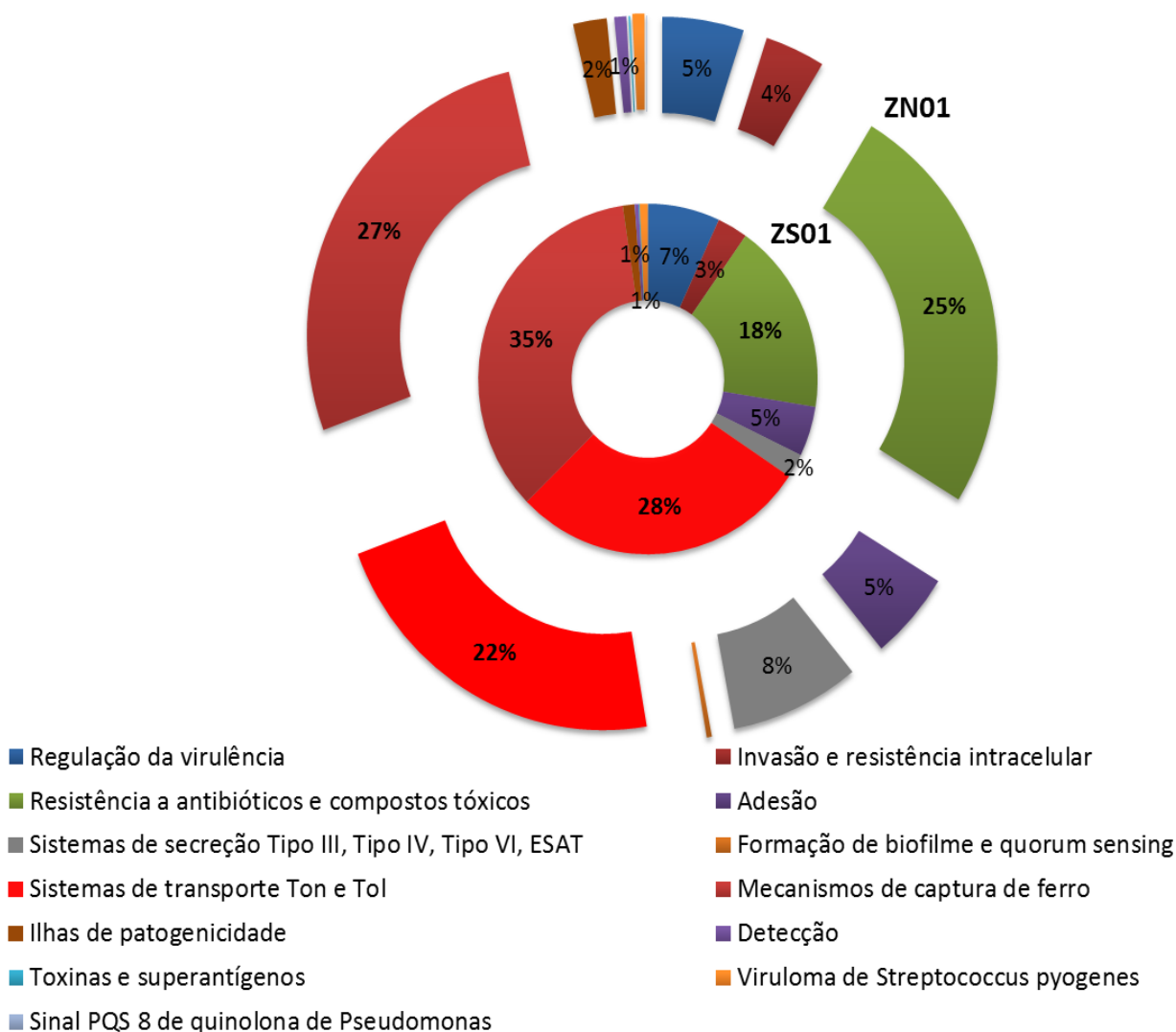
**Figura 5.35. Classificação funcional dos subsistemas presentes nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) utilizando a base SEED através do programa MEGAN 4. O gráfico indica a abundância relativa de cada grupo funcional (subsistema) relacionado às bactérias presentes nas amostras ZS01 e ZN01. “Not assigned” são as sequências que tiveram uma correspondência com a base de dados RefSeq de proteína que não foram atribuídas a um táxon. “No hits” indica as sequências que não apresentaram nenhuma correspondência pelo resultado do BLASTX contra a base RefSeq de proteína.**

Em relação ao metabolismo de carboidratos, os subsistemas mais prevalentes, em ambas as amostras ZS01 e ZN01 foram respectivamente: o de monossacarídeos (28 e 23%), o de Di- e oligossacarídeos (21 e 23,3%) e o de metabolismo central de carboidratos (13 e 18%) (Figura 5.36).



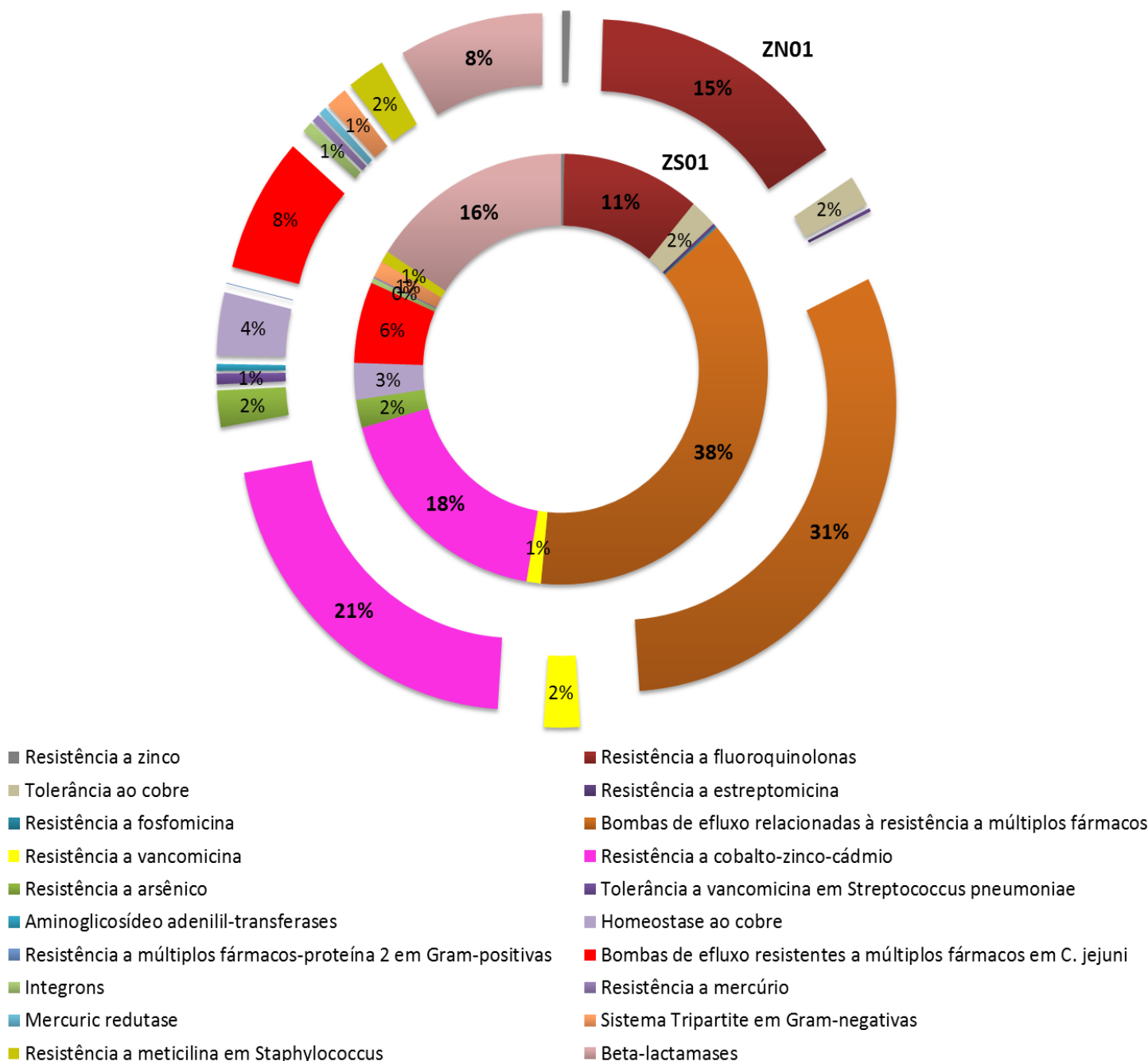
**Figura 5.36. Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de carboidratos presente nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) pela base SEED presente no MEGAN4.**

Dentre as proteínas relacionadas ao subsistema de virulência, as predominantes em ZS01 e ZN01 foram respectivamente: mecanismos de captação de ferro (Fe<sup>+2</sup>) (35 e 27%), sistema de transporte de Ton e Tol (29 e 22%), resistência a antibióticos e compostos tóxicos (18 e 25%) e regulação de virulência (7 e 5%) (Figura 5.37).



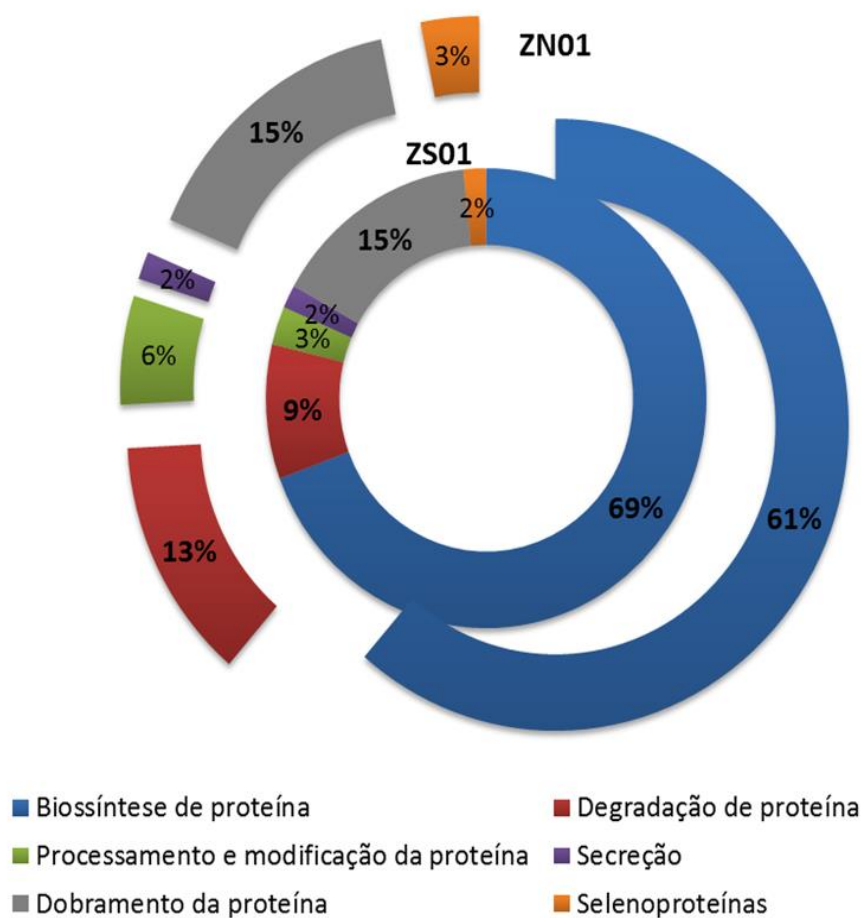
**Figura 5.37. Classificação funcional das proteínas relacionadas ao subsistema de virulência presente nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS, círculo interno) e Zona Norte (ZN, círculo externo) pela base SEED presente no MEGAN4.**

Em relação à classificação das proteínas relacionadas com a resistência a antibióticos e aos compostos tóxicos, os subsistemas que predominaram em ambas as amostras ZS01 e ZN01 foram respectivamente: sistema de bombas de efluxo ligadas à multiresistência (38 e 31,3%), resistência a cobalto-zinco-cádmio (17 e 21%), resistência às fluoroquinolonas (11 e 15,5%) e  $\beta$ -lactamases (16% e 8%) (Figura 5.38).



**Figura 5.38. Classificação funcional das proteínas relacionadas ao subsistema de resistência a antibióticos e compostos tóxicos presentes nas réplicas 01 das amostras dos hospitais da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo) pela base SEED presente no MEGAN4.**

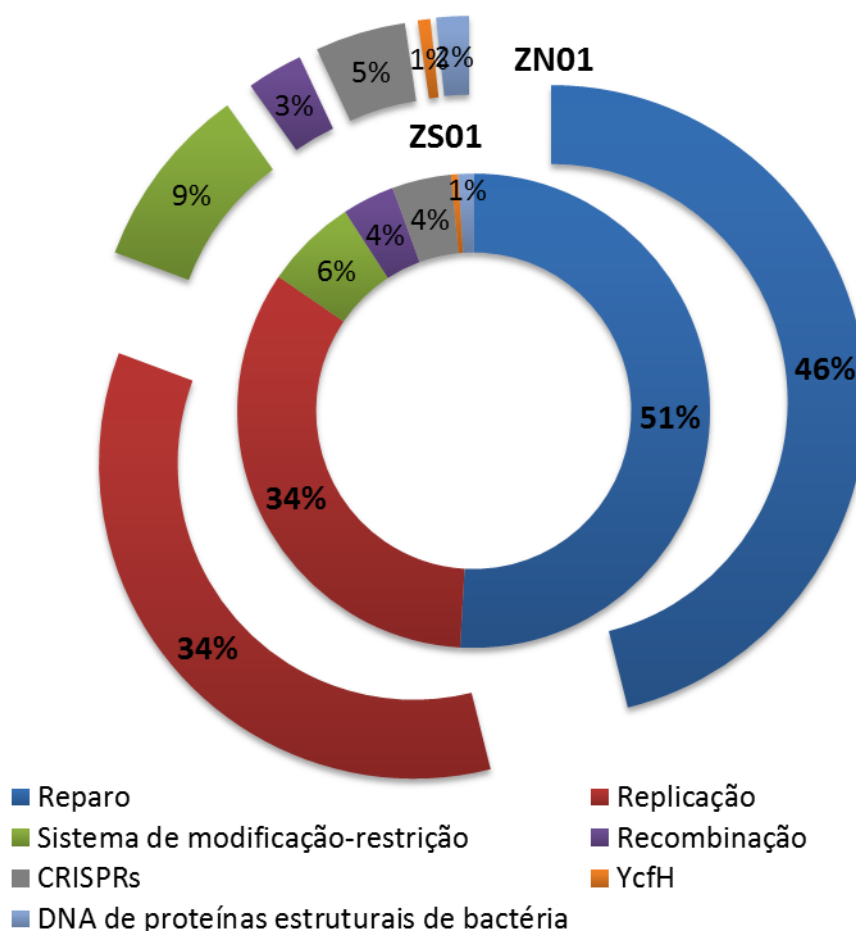
Em relação ao subsistema de metabolismo de proteínas predominaram as proteínas relacionadas à biossíntese de proteína com 70 e 61%, para as amostras ZS01 e ZN01, e ao dobramento de proteínas com 15 e 15,4% para ZS01 e ZN01, respectivamente (Figura 5.39).



**Figura 5.39. Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de proteínas presentes nas réplicas 01 das amostras hospitalares da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo), pela base SEED presente no MEGAN4.**



Considerando o subsistema de metabolismo de DNA as proteínas mais abundantes foram às relacionadas ao reparo, com 51 e 46% para as amostras ZS01 e ZN01, respectivamente. Em segundo lugar foram proteínas do subsistema de replicação com, 34 e 35%, para ZS01 e ZN01, respectivamente (Figura 5.40).



**Figura 5.40. Classificação funcional das proteínas relacionadas ao subsistema de metabolismo de DNA presente nas réplicas 01 das amostras hospitalares da Zona Sul (ZS01, círculo interno) e Zona Norte (ZN01, círculo externo), pela base SEED presente no MEGAN4.**

## 6 DISCUSSÃO

### PARTE I: Parte experimental

#### 6.1 Parâmetros físico-químicos e colimetria das amostras de esgotos hospitalares do Rio de Janeiro

A grande diversidade de atividades de um hospital público pode gerar um efluente bastante complexo, cuja combinação de substâncias e de micro-organismos pode causar efeitos praticamente desconhecidos ao meio ambiente, mas de grande interesse para estudos de saúde pública. Isso destaca o papel da água na circulação de micro-organismos resistentes aos antibióticos no meio ambiente (Vecchia *et al.*, 2009).

Em relação aos parâmetros físico-químicos quantificados nesse trabalho, algumas considerações podem ser colocadas. A DBO avalia a quantidade de oxigênio dissolvido (OD) que será consumida na oxidação biológica da matéria orgânica. A partir desta se estima a carga orgânica de corpos d'água, de esgotos, e a necessidade de aeração para degradá-la em estações de tratamento de esgoto (Grenberg *et al.*, 2005). Segundo Von Sperling (2009) o elevado valor de DBO sugere poluição orgânica que acaba sendo degradada por micro-organismos, acarretando no decréscimo de oxigênio dissolvido na água. Verlicchi e colaboradores (2010) observaram que este parâmetro pode variar entre 90 a 200 mg/L de O<sub>2</sub>, dependendo do tamanho da unidade de esgotos não tratados. Segundo Paz e colaboradores (2004), em seu trabalho de caracterização físico-química de esgotos hospitalares, este parâmetro não deve ultrapassar os 200mg/L. Os esgotos dos hospitais da Zona Norte e Sul possuem valores mais elevados (295 mg/L e 209 mg/L, respectivamente), ultrapassando o valor recomendado.

Em relação ao DQO, Von Sperling (2009) menciona que este parâmetro também se relaciona com a concentração de oxigênio utilizado para consumir a matéria orgânica, biodegradável ou não, em meio ácido e condições energéticas por ação de um agente químico oxidante forte. Em relação a esse parâmetro, Verlicchi e colaboradores (2010) citam que esgotos não tratados de unidades de saúde podem variar entre 170 e 500 mg/L de O<sub>2</sub>, dependendo da

característica do efluente e tamanho da unidade de saúde. Paz e colaboradores (2004) relatam que este componente não deve ultrapassar 700mg/L, ou seja, ambos os esgotos dos dois hospitais se encontram com o valor de DQO dentro da recomendação (Tabela 5.1).

O fósforo total nos esgotos sanitários aparece nos compostos orgânicos (proteínas) ou compostos minerais (polifosfatos e ortofosfatos) que se acredita virem, principalmente, de despejos de detergentes sintéticos e produtos utilizados para a assepsia do local. Quando descartados em alta concentração nos leitos, junto com nitrogênio total, são nutrientes para micro-organismos (Evaldt, 2005). Segundo a resolução do CONAMA, a concentração ideal de fósforo total em um efluente, ao ser descartado, deve ser de até 1 mg/L. Os dois hospitais públicos apresentaram valores bem mais elevados do que o recomendado (12 e 14 mg/L para ZS e ZN, respectivamente). Estes resultados estão em acordo com outros valores encontrados em trabalhos de caracterização físico-química de esgotos hospitalares da cidade de Porto Alegre (Evaldt, 2005; Ribeiro, 2005).

Em esgotos sanitários, o nitrogênio total normalmente é encontrado na forma de carbonato de amônia, resultado da hidrólise da uréia presente na urina. Os nitrogênios orgânicos são encontrados nas moléculas de proteínas ou de aminoácidos ainda não assimiladas (Von Sperling, 1998; Evaldt, 2005). Segundo a resolução do CONAMA, a concentração ideal de nitrogênio total em um efluente, ao ser descartado, deve ser de até 10mg/L. Os dois hospitais apresentaram novamente valores bem mais elevados do que o recomendado (55,3 e 55 mg /L para ZS e ZN, respectivamente). Os valores mais elevados que a recomendação torna preocupante o descarte deste efluente se o mesmo não tiver um tratamento prévio (Evaldt, 2005). Desta forma, é importante que estes esgotos hospitalares sejam tratados antes de serem descartados diretamente nos corpos de água.

A contagem de coliformes fecais, como era de se esperar, apresentou valores bem elevados, uma vez que esse efluente é de esgoto “in natura”, mostrando uma contaminação fecal recente. A contagem de coliformes totais também foi muito elevada, ultrapassando o crescimento que permitisse a contagem na última diluição utilizada ( $10^{-3}$ ).

## **6.2 Informações sobre os dados gerados pelo pirosequenciamento das amostras de EHRJ**

O rendimento e o desempenho do sequenciamento pelo sistema 454 GS FLX + foi satisfatório, dentro do que a Roche propõe para essa tecnologia. Esse rendimento e tamanho das sequências geradas é superior às tecnologias anteriores do 454/Life Science (como 454 FLX e ainda o 454 GS FLX Titanium) pois gera um tamanho de sequências em torno de 600-1000pb. Essa nova tecnologia facilita muito a análise das sequências em projetos de metagenoma, aprimorando e elevando o rendimento da classificação correta destas.

## **PARTE 2: Parte “In silico” (Não experimental)**

### **6.3 Caracterização do metagenoma dos esgotos hospitalares do Rio de Janeiro**

#### **6.3.1 *Perfil taxonômico da comunidade bacteriana presente nas amostras de esgotos dos hospitais da Zona Norte e Zona Sul***

A composição dos esgotos hospitalares é variável e se relaciona com as atividades que o hospital desempenha. Normalmente eles são compostos por dejetos humanos, os quais, por estarem em um ambiente propício ao desenvolvimento de doenças, possuem um grande número de microorganismos patogênicos e resistentes a múltiplas drogas (Ribeiro, 2005). Está incluída também nestes esgotos a água de lavagem de materiais contaminados, os dejetos de limpeza de superfícies e pisos misturados a soluções desinfetantes, a água da lavanderia (quando há), as águas das caldeiras, os resíduos de procedimentos em centros cirúrgicos, dos ambulatórios, dos laboratórios de análises clínicas e anatomopatológicos (Ribeiro, 2005). Porém, vale ressaltar que os dejetos de indivíduos internados em UTIs (Unidade de Tratamento Intensivo) não vão para os esgotos dos hospitais, mas para um descarte especial, os descartes de resíduos hospitalares.

Sabe-se que o trato gastrointestinal de um humano adulto contém os três domínios da vida, Bacteria, Archaea e Eukarya, além de vírus. Porém, a

maioria do seu microbioma é composta por bactérias que vivem no intestino humano. Elas atingem as maiores densidades de célula já registradas para qualquer ecossistema e estima-se que 800-1000 diferentes espécies de bactérias habitam o trato intestinal (Bäckhed *et al.*, 2005). Pouco se sabe, ainda, sobre essa vasta comunidade microbiana. Estudos recentes revelaram que a maioria da microbiota (80%) do intestino humano ainda não foi cultivada (Schmieder & Edwards, 2012). Após estudos com diversos indivíduos, foi definido um “core” do microbioma intestinal humano, composto pelos grupos de linhagens presentes na metade dos indivíduos envolvidos no estudo (Qin *et al.*, 2010). Os principais grupos que compõem esse “core” de diversidade de Bacteria, e logo predominam nas fezes humanas, são: Bacteroides (sendo o gênero *Bacteroides* o mais abundante) e Firmicutes, sendo 95% das sequências pertencentes à classe Clostridia (os gêneros mais prevalentes são *Clostridium*, *Ruminococcus* e *Eubacterium*) (Eckburg *et al.*, 2005; Gil *et al.*, 2006). Esses dois grupos juntos compõem mais de 90% da microbiota intestinal humana (Eckburg *et al.*, 2005). O grupo Proteobacteria embora comum, não é dominante. O microbioma intestinal difere bastante em relação à idade da pessoa, como entre um neonatal e um adulto, pois ele está relacionado com o tipo de alimentação e o sistema imunológico do indivíduo (Kurokawa *et al.*, 2007; Palmer *et al.*, 2007; Ursell *et al.*, 2012). Espera-se que a microbiota intestinal humana tenha uma contribuição significativa no metagenoma do esgoto hospitalar embora outros fatores também devam ser considerados. Além disso, vários fatores podem contribuir para variações no metagenoma de esgoto hospitalar e precisam ser considerados.

Comparando com trabalhos da literatura, os resultados aqui discutidos apresentam variações em relação à diversidade taxonômica, pois utilizam diferentes plataformas de sequenciamento envolvendo diferentes metodologias com suas diferentes limitações. Em relação ao perfil taxonômico de Bacteria, a partir dos dados gerados pelo pirosequenciamento, pôde-se observar que a riqueza e diversidade da comunidade microbiana presente na amostra de esgoto hospitalar da Zona Norte (ZN) do Rio foi maior que a da Zona Sul (ZS) (Figura 5.14). Essa diferença pode ser devido ao fato de que o local de coleta da amostra do hospital da Zona Sul não abrangeu todos os setores, mas apenas quatro (cardiologia, ortopedia, laboratório de análises clínicas e clínica médica). Já os esgotos coletados do hospital da Zona Norte (ZN) abrangiam

todos os setores do hospital. Outro possível fator que pode ter contribuído para essa diferença pode ser a quantidade de material químico utilizado na lavagem dos banheiros e outros setores e a frequência com a qual esse material é utilizado. Porém, como não foram realizados testes para a detecção e quantificação de detergentes e outros compostos químicos utilizados na limpeza dos hospitais não é possível confirmar essa suposição. Vale a pena ressaltar também que o hospital ZN possui uma maior complexidade, com diversas especializações, consultas e tratamentos, no qual o paciente fica internado por um período mais longo. Já no hospital ZS, que é um hospital de emergência, os tratamentos são mais rápidos e a permanência dos pacientes é por um período curto.

A diferença mais acentuada da distribuição da comunidade bacteriana entre os dois hospitais foi em relação à Proteobacteria (Figura 5.28), principalmente em relação à classe Gammaproteobacteria (Figura 5.31). Essa diferença pode estar relacionada a um maior descarte de antibióticos pelo hospital da Zona Norte do Rio no momento da coleta (consequentemente o efluente apresentaria uma maior quantidade de antibióticos), ou devido à diferença da região de coleta. Sabe-se que a terapia de antibióticos pode afetar não só o patógeno alvo, mas também os comensais hospedeiros intestinais humanos (Jernberg *et al.*, 2010). Kristiansson e colaboradores (2011) verificaram que, em sedimentos de uma estação de tratamento de esgoto, com altas cargas de antibióticos, houve predominância de bactérias pertencentes aos filos Firmicutes, Bacteroidetes e Proteobacteria. Segundo Mariat e colaboradores (2009) quando há tratamento com antibióticos as bactérias pertencentes ao grupo Proteobacteria se tornam mais abundantes, sugerindo que possuem um aumento relativo em seu desempenho (“fitness”) durante o tratamento com antibióticos (Sommer *et al.*, 2010).

Até o presente não foi confirmada a existência de trabalhos publicados e acessíveis que analisaram a diversidade de Bacteria em esgotos hospitalares utilizando a tecnologia de pirosequenciamento através do “Shotgun”, mas apenas de trabalhos relacionados com estação de tratamento de esgoto (Schlüter *et al.*, 2008; Sanapareddy *et al.*, 2009; Kristiansson *et al.*, 2011). Os poucos trabalhos relacionados à diversidade microbiana de esgotos hospitalares são de sequências de 16S ribossomal (rrs) obtidas pelo método de Sanger (Chagas *et al.*, 2011). A maioria dos trabalhos de metagenoma

utilizando sequenciamento em larga escala que envolve diversidade de Bacteria trabalha com análise de “amplicons” de genes marcadores, como a região 16S ribossomal. Geralmente, os estudos que envolvem apenas análise de diversidade não utilizam a abordagem de “Shotgun”. Sendo assim, os dados utilizados para as comparações com os resultados obtidos no presente trabalho incluem, também, estudos de estações de tratamento de esgoto, utilizando o gene do 16S ribossomal, além do sequenciamento por tecnologias em larga escala (como o 454/Life Science e Illumina/Solexa). Além disso, também foram utilizados dados de trabalhos que analisaram a diversidade do microbioma intestinal humano já que as amostras coletadas e analisadas continham, em sua maioria, fezes humanas. Vale ressaltar que as condições ambientais do esgoto hospitalar são muito diferentes do intestino humano. Mesmo assim, os resultados de composição taxonômica foram muito semelhantes, principalmente em relação ao filo.

Chagas e colaboradores (2011) estudaram a diversidade da comunidade bacteriana de um efluente hospitalar localizado na Zona Oeste do Rio de Janeiro (“in natura” e tratado). Eles observaram, na amostra “in natura”, uma maior abundância de Proteobacteria, Firmicutes e Spirochaetes. Esse resultado difere bastante do encontrado no presente estudo, no qual o grupo Spirochaetes foi encontrado em uma proporção muito baixa (0,13 e 0,08% para ZN e ZS, respectivamente). A metodologia utilizada se baseou na análise de sequências parciais da região 16S ribossomal. Porém, a análise da comunidade de Bacteria através do 16S ribossomal realizada nesta tese, revelou resultados semelhantes aos obtidos pela metodologia de pirosequenciamento. Logo, a distribuição taxonômica da comunidade parece ser realmente distinta entre o esgoto do hospital da zona Oeste e os outros dois analisados neste estudo (ZS e ZN).

Em relação à diversidade de Bacteria, os resultados encontrados no presente trabalho foram semelhantes aos descritos nos trabalhos anteriormente mencionados, ou seja, houve a predominância do grupo Firmicutes (53-56%), que são Gram-positivas anaeróbias na sua maioria, e Bacteroidetes (27-33%), que são Gram-negativas anaeróbias estritas na sua maioria.

Em relação aos gêneros encontrados pertencentes aos filios Firmicutes e Bacteroidetes, os resultados desse trabalho são semelhantes aos encontrados por Palmer e colaboradores (2007), assim como o trabalho de

Mariat e colaboradores (2009). A maioria das sequências classificadas foi associada a gêneros que compõem a comunidade de Bacteria do intestino humano de adultos saudáveis como os anaeróbios *Bacteroides*, *Eubacterium*, *Clostridium*, *Ruminococcus*, *Faecalibacterium*, *Blautia* e *Coprococcus*. A diferença foi em relação à *Clostridium*, que não foi abundante nas amostras de esgotos hospitalares, mas sim o gênero *Eubacterium*. Como as condições ambientais dos esgotos hospitalares e intestino humano são muito distintas, não se sabe se essa comunidade bacteriana detectada nos esgotos hospitalares do Rio foi capaz de sobreviver a essas condições ambientais ou se parte não sobreviveu e apenas seu material genético (DNA) estava íntegro.

#### **6.4 Diversidade de $\beta$ -lactamases em esgotos hospitalares e em projetos de metagenomas públicos**

É importante ressaltar que os projetos dispostos nos bancos públicos CAMERA e IMG/M possuem tamanhos distintos gerados por diferentes tecnologias como pelo método de sequenciamento de ABI Sanger (plataforma ABI3730XL), pela plataforma 454/Life Science e a plataforma Illumina/Solexa. Contudo, todas as sequências utilizadas do IMG/M estavam em forma de “contigs”, ou seja, montadas por sobreposição, enquanto as sequências utilizadas do CAMERA estavam “cruas” (apenas foram filtradas por qualidade, redundância e artefatos). Sabe-se que o método de Sanger gera sequências entre 500-1000 pares de bases e possui o viés da clonagem, enquanto o 454/Life Science, utilizado nos projetos metagenômicos minerados, gera sequências de tamanhos menores (~ 250-500 pb), e o tamanho médio das  $\beta$ -lactamases varia de 250 a 400 aminoácidos (~750-1200 nucleotídeos). Desta forma, realizou-se um estudo voltado para a exploração da distribuição dessas enzimas nos esgotos e nos projetos metagenômicos relacionados a diferentes ambientes, além de verificar a relação filogenética entre elas e as  $\beta$ -lactamases produzidas por patógenos clínicos, assim como classificá-las. Sendo assim, de acordo com o resultado obtido no item 5.8.1 (Figura 5.6) as  $\beta$ -lactamases das classes A e C apresentaram frequências semelhantes em cada amostra hospitalar (ZS e ZN). Contudo, a classe D apresentou uma frequência bem mais elevada na amostra ZN (~80%) em relação à ZS (~20%). Como o efluente da Zona Norte abrange todos os locais do hospital e os coletados no hospital



da Zona Sul abrangem apenas alguns setores (cardiologia, clínica médica, ortopedia e laboratório de análises clínicas) pode ser que as  $\beta$ -lactamases da classe D possam ser encontradas com mais frequência em certos setores que não foram contemplados nas amostras coletadas no hospital da Zona Sul devido ao uso diferenciado de classes de  $\beta$ -lactâmicos nos diferentes setores. Esta hipótese poderia ser testada pelo cruzamento de dados sobre as condições dos pacientes e tipos de doenças que circulam nesses hospitais, além do uso de cada classe de antibióticos nesses diferentes setores. Porém, esses dados não estão livremente disponíveis. É possível que nos setores que não foram contemplados no hospital da Zona Sul ocorra um uso maior de  $\beta$ -lactâmicos como, por exemplo, cloxacilina, oxacilina e meticilina.

Nos resultados de abundância relativa de  $\beta$ -lactamases das bases metagenômicas (CAMERA e IMG/M) (item 5.8.2) as três classes de serina  $\beta$ -lactamases apresentaram taxas de frequência totalmente distintas nos diferentes projetos. Os projetos que apresentaram maior frequência da classe A foram o de Mina ácida (“Acide Mine”) e o de metagenoma de “Yellowstone Bison Hot Spring” (BISON). Essa elevada abundância se refletiu na árvore filogenética (Figura 5.12) que também apresentou muitas sequências oriundas de ambos os projetos, formando um grande clado. Uma possível explicação para esse fato pode ser a média do comprimento das sequências, que foram maiores nesses dois projetos (> 900 pb) aumentando assim a chance de terem sequências inteiras de  $\beta$ -lactamases e serem recuperadas pelo programa HMMER. Por isso, mesmo após a busca com o “hmmsearch” utilizando os pHMMs das três regiões principais da classe A, a maioria das sequências resultantes foram desses dois projetos que são de ambientes extremos. De qualquer forma, os projetos GOS (“Global Ocean Sampling Expedition”) e o de intestino de humano (“Human Gut”) também possuem essas mesmas características (tamanho de sequências maiores), mas, no entanto as  $\beta$ -lactamases classe A não se mostraram tão frequentes quanto nesses projetos. Entretanto, vale ressaltar que esse resultado é relativo e tendencioso, pois como existem muitas sequências de  $\beta$ -lactamases classe A mineradas do projeto de Mina ácida (“Acid Mine”), a relação das classes A nos outros projetos parece muito menor e, em alguns, inexistentes, enquanto em projetos como “Human Distal Gut” o valor absoluto é o mesmo. Em relação ao projeto metagenômico “Yellowstone Bison Hot Spring (BISON)”, foram coletadas

amostras de cinco locais diferentes ao longo de um canal situado no parque de Yellowstone, que possui fontes termais. Curiosamente, as  $\beta$ -lactamases classe A recuperadas foram apenas nos sítios 3 e 4 desse projeto. O sítio 3 apresenta um ambiente mais diverso e hostil, além de possuir uma rigorosa limitação de nutrientes. O sítio 4 é o que possui a temperatura mais alta ( $> 68^{\circ}\text{C}$ ) (Swingley, *et al.*, 2010).

Em relação às  $\beta$ -lactamases classe C, estas foram mais abundantes em projetos de ambientes distintos como aquático (“DeepMed”) e de solo (“Farm Soil”). O primeiro projeto é da comunidade microbiana de regiões profundas do mar Mediterrâneo, também conhecidas como zona batipelágica, que é uma camada da zona pelágica que se estende da profundidade de 1000m a 4000m abaixo do oceano. São regiões aonde a luz não alcança, possuem baixa temperatura e uma pressão cada vez menor com o aumento da profundidade (Martin-Cuadrado *et al.*, 2007). Já o projeto “Farm Soil” consiste na comunidade microbiana da superfície (0-10cm) de um solo de região agrícola, em Minnesota (EUA) (Tringe *et al.*, 2005). Em relação aos projetos nos quais as  $\beta$ -lactamases classe D foram mais abundantes destacaram-se o “Human Distal Gut”, da microbiota intestinal humana e “PBSM” que é da comunidade microbiana da areia das praias do Pacífico, ambos de ambientes distintos também. Esses resultados sugerem uma heterogeneidade das diferentes classes de  $\beta$ -lactamases nos diferentes ambientes e, provavelmente, uma grande dispersão destas, mesmo em ambientes onde não há (ou há pouca) presença de antibióticos. No entanto, estudos de elementos genéticos móveis presentes nesses projetos podem ser analisados de forma ajudar a contribuir para a suposição da mobilidade e dispersão destas enzimas.

Os pHMMs construídos foram capazes de recuperar prováveis serina  $\beta$ -lactamases a partir de sequências representativas, sendo a maioria curada, em diferentes ambientes, mostrando ser uma ferramenta adequada para o estudo do resistoma de genes de resistência a  $\beta$ -lactâmicos, contribuindo para a classificação das diferentes classes (A, C e D). A base IMG/M (Markowitz *et al.*, 2007) é anotada de forma automática através de comparações de busca por motivos e domínios, através do RPS-BLAST, utilizando bases como COG (Tatusov *et al.*, 1997) e Pfam (Bateman *et al.*, 2004). Os pHMMs foram capazes de buscar possíveis  $\beta$ -lactamases descritas como “proteínas hipotéticas” na anotação do IMG/M, confirmadas depois pelas análises do

RPS-BLAST e filogenia, na qual além de apresentarem anotação de  $\beta$ -lactamases, agruparam próximas às sequências de referência utilizadas na análise filogenética.

Bactérias comensais do intestino (microbiota) são conhecidas por abrigar genes de resistência e, assim, contribuir para o reservatório de genes de resistência no intestino que podem ser adquiridos por membros patogênicos da comunidade intestinal (Sommer *et al.*, 2009). Segundo Sommer e colaboradores (2009), que fizeram um estudo do resistoma presente no microbioma intestinal humano de indivíduos saudáveis e que estavam sem o uso de antibióticos por um ano, verificaram que os genes de resistência encontrados nos indivíduos que administraram o antibiótico eram distintos e distantes evolutivamente dos genes de resistência conhecidos atualmente. Um resultado semelhante foi observado no presente estudo. A maioria dos prováveis homólogos de  $\beta$ -lactamases oriundas das amostras de esgotos hospitalares do Rio de Janeiro não se agrupou próxima às  $\beta$ -lactamases utilizadas como representativas de bactérias pertencentes ao grupo Proteobacteria nas árvores filogenéticas. Essas enzimas se agruparam próximas de sequências representativas de Firmicutes (nas classes C e D) (Figura 5.13 e 5.14) e Bacteroidetes (na classe A), principalmente, (Figura 5.11), incluindo gêneros como *Bacteroides*, *Prevotella*, *Roseburia*, *Eubacterium*, *Ruminococcus* e *Clostridium*.

Já as sequências de  $\beta$ -lactamases recuperadas de projetos ambientais do CAMERA e IMG/M se agruparam mais próximas às  $\beta$ -lactamases representativas do filo Proteobacteria (como Gammaproteobacteria) e algumas sequências de Alphaproteobacteria e Epsilon-Proteobacteria, nos casos das filogenias das classes C e D (Figuras 5.13 e 5.14). Essas prováveis  $\beta$ -lactamases eram de projetos distintos, como de um ambiente aquático (mar de Sargasso), projetos de estação de tratamento de esgoto, da comunidade microbiana de artrópodes que possuem uma interação simbiótica e de projetos de solo (comunidade microbiana de solos) (Figuras 5.12, 5.13, 5.14).

Na filogenia relacionada às prováveis  $\beta$ -lactamases classe A, dos esgotos hospitalares se observou que a maioria das sequências pertence ao grupo Bacteroidetes, o segundo grupo taxonômico mais abundante das quatro amostras (27-33%), sendo a maioria pertencente ao gênero *Bacteroides* (~46%) (incluindo *B. vulgatus* e *B. fragilis*, que são oportunistas), da família

Bacteroidaceae. Esses micro-organismos são, em sua maioria, comensais da microbiota intestinal humana, Gram-negativos e anaeróbios estritos. Na filogenia da classe D a maioria das sequências das amostras hospitalares se agruparam próximas de sequências pertencentes ao filo Firmicutes, enquanto que na filogenia da classe C as sequências se agruparam próximas a gêneros de ambos os grupos.

A análise metagenômica da microbiota intestinal mostrou uma inesperada riqueza de diferentes  $\beta$ -lactamases, sendo a maioria delas desconhecidas em bactérias patogênicas cultiváveis. Sommer e colaboradores (2009) verificaram que a maioria dos genes de resistência a antibióticos pertencentes ao microbioma intestinal de humanos apresentou uma relação evolutiva distante (60,7% em nível de nucleotídeo e 54,9% em nível de aminoácidos) dos genes de resistência (e seus produtos) detectados, até o presente, em isolados patogênicos. Na análise filogenética que realizaram com esses genes verificaram que a maioria originou dos filios Bacteroidetes e Firmicutes, mas mesmo assim esses genes sequenciados apresentaram uma baixa similaridade com sequências de patógenos pertencentes a esses filios. Quando esse grupo verificou as sequências do microbioma intestinal de dois humanos adultos, os quais haviam administrado antibiótico, eles perceberam que a comunidade microbiana não só mudou, com um aumento de Proteobacteria e diminuição de Bacteroidetes e Firmicutes, mas também que a maioria (>95%) dos genes de resistência aos antibióticos sequenciados das fezes desses indivíduos apresentou uma alta similaridade com os genes de resistência de patógenos clínicos. Sommer e colaboradores (2009) analisaram o microbioma intestinal humano e verificaram uma riqueza inesperada de diferentes  $\beta$ -lactamases, sendo a maioria desconhecida em cultura de bactérias patogênicas.

Em um trabalho de 2010, esses mesmo autores identificaram 10 novas sequências de famílias de  $\beta$ -lactamases (HGA até HGI e HOA) que apresentaram uma identidade abaixo de 61%, em nível de aminoácidos, em relação às sequências previamente caracterizadas. Baseado em análises filogenéticas, essas enzimas mostraram ser produzidas predominantemente por bactérias do filo Bacteroidetes e Firmicutes. Esses resultados indicam que os genes de resistência a antibióticos no microbioma intestinal humano está sub-representado, uma vez que a maior parte dos dados vem dos micro-

organismos cultiváveis. Sendo assim, a metagenômica aparece como uma abordagem útil e importante para o estudo dos reservatórios de genes de resistência a antibióticos.

Inicialmente as sequências curadas de  $\beta$ -lactamases utilizadas nas construções dos pHMM foram inseridas nas filogenias como as sequências representativas para a classificação das prováveis  $\beta$ -lactamases mineradas pelo “hmmsearch” a partir dos pHMMs de cada classe. Porém, essas sequências não se agruparam com nenhuma sequência de  $\beta$ -lactamase das amostras hospitalares. Esse fato deve ter ocorrido porque essas sequências curadas são de bactérias pertencentes ao filo Gammaproteobacteria, sendo a maioria patogênica e isolada de áreas clínicas. Como o resultado da distribuição taxonômica mostrou que a abundância das bactérias foi dos filos Bacteroidetes e Firmicutes, entende-se a razão pela qual esse agrupamento na filogenia não tenha ocorrido (são bactérias pertencentes a filos diferentes). Desta forma, tornou-se necessário a busca de sequências representativas que pertencessem a esses filos, por isso foi realizado um BLASTP contra o Refseq de proteínas utilizando as sequências de esgotos hospitalares obtidas pelo “hmmsearch” como entrada. Como essas sequências não são curadas manualmente, a análise por similaridade contra a base de perfil PSSM do COG foi realizado para conferir um maior suporte na anotação da função dessas sequências (conforme descrito no item 4.8).

Poucas sequências homólogas de  $\beta$ -lactamases dos dados de esgotos hospitalares foram recuperadas (Tabela 5.6), pois, apesar do sequenciamento ter apresentado um rendimento excelente gerando tamanhos de 700-900 pares de base, foram geradas sequências de aproximadamente 150-300 aminoácidos, e as  $\beta$ -lactamases possuem em torno de 250-400 aminoácidos. Isso explica o fato de muitas sequências não apresentarem a região conservada Lys-Thr (Ser)-Gly [K-T(S)-G] localizada na porção final da sequência primária da enzima próximo ao N-terminal. Como essas sequências possuíam as demais regiões conservadas das classes e apresentaram anotação de  $\beta$ -lactamases a partir de análise por similaridade com bases de dados de domínios conservados (RPS-BLAST), com valores de “E-values” conservadores (abaixo de  $10^{-20}$ ), foram utilizadas na reconstrução da filogenia.

Em relação às prováveis  $\beta$ -lactamases classe A detectadas nos projetos metagenômicos (que não o de EHRJ), as subclasses CTX-M e TEM foram

encontradas em diferentes ambientes, incluindo aqueles aonde existe pouca atividade antropogênica e, aparentemente não há muita presença de antibióticos, como o mar de Sargasso e as águas termais de Yellowstone. Essa observação sugere uma mobilidade dessas enzimas no meio ambiente, revelando um reservatório de genes de resistências a antibióticos  $\beta$ -lactâmicos. Em estudos anteriores já havia sido relatada essa disseminação dos genes de resistência, incluindo as  $\beta$ -lactamases presentes em solo como da Antártica, que não tinha nenhum tipo de contaminação por antibióticos (Allen *et al.*, 2009). Por outro lado, estudos apontam que os antibióticos, incluindo as  $\beta$ -lactamases, podem agir como moléculas sinalizadoras e modular a expressão gênica bacteriana em concentrações subinibitórias (Goh *et al.*, 2002; Yim *et al.*, 2007).

Em relação às prováveis subclasses de  $\beta$ -lactamases classe A detectadas nas amostras ZS e ZN, se observou uma maior abundância de  $\beta$ -lactamases CfxA (65% para ZS e 28,4% para ZN) e CEPA (17% para ZS e 20% para ZN) (Figura 5.7). Segundo alguns trabalhos (Ferreira *et al.*, 2007; Tribble *et al.*, 1999) a transferência horizontal de genes de  $\beta$ -lactamase é frequente entre espécies de *Bacteroides*, inclusive para o gene *cfxA* que geralmente está associado com um transposon conjugativo (Tn4555). A presença desse gene pode levar a uma resistência a altos níveis de cefoxitina (Tribble *et al.*, 1997). Esse gene foi inicialmente descrito em *B. vulgatus* (Rogers *et al.*, 1993). Essa elevada abundância de CfxA pode estar relacionada com a abundância de *Bacteroides* encontrados nas amostras de esgotos hospitalares, sugerindo a possibilidade desse gene estar disseminado entre as espécies de *Bacteroides* comensais do intestino humano, indicando um possível reservatório de genes de resistência à antibióticos (no caso dos  $\beta$ -lactâmicos, principalmente cefalosporinas).

Outra provável  $\beta$ -lactamase classe A detectada principalmente nas amostras da ZN, apresentando uma abundância relativa de aproximadamente 10% em relação ao total de  $\beta$ -lactamases classe A classificadas (Figura 5.5) foi a KPC (*K. pneumoniae* carbapenemase). Essas enzimas são carbapenemases, ou seja, hidrolisam carbapenemas que são os antibióticos  $\beta$ -lactâmicos mais eficazes até o momento, sendo a maioria encontrada em plasmídeos. Até o momento são conhecidas 10 variantes de KPC (KPC-2 até KPC-11). Essas enzimas já foram descritas também em outras espécies de

Enterobacteriaceae (Cuzon & Nordmann, 2010), incluindo em isolados do Brasil (Peirano *et al.*, 2009), sendo primeiro descrita no país em 2007 (Monteiro *et al.*, 2007), apesar de ter sido isolada e descrita primeiro nos EUA em 1996 (Yigit *et al.*, 2001). As subclasses de  $\beta$ -lactamases ESBLs como as PER, VEB, GES e IBC foram descritas, porém não são tão comuns, encontradas basicamente em *P. aeruginosa* e em um número pequeno de regiões geográficas como França e Turquia (PER) (Vahaboglu *et al.*, 1997; De Champs *et al.*, 2002; ), França e África do Sul (GES) (Poirel *et al.*, 2002; Dubois *et al.*, 2002), Tailândia, Kwait e China (VEB) (Girlich, *et al.*, 2001; Naas *et al.*, 1999; Tribuddharat & Fennewald, 1999; Jiang *et al.*, 2005; Poirel *et al.*, 2001). No presente trabalho foram detectadas prováveis  $\beta$ -lactamases GES (capaz de hidrolisar carbapenêmicos), VEB e PER (classe A) (Figura 5.7). Até o momento, no Brasil, essas subclasses têm sido descritas em menor número, com pouquíssimos relatos (Silva & Lincopan, 2012). No entanto, essas três subclasses são ESBLs, o que é preocupante para a saúde pública. Variantes das ESBLs do tipo GES são relatadas, principalmente, em isolados de *P. aeruginosa* recuperados de pacientes internados em hospitais (Castanheira *et al.*, 2004; Pellegrino *et al.*, 2006), porém variantes dessa subclasse já foram detectadas em *K. pneumoniae* no Estado de São Paulo (Dropa, 2010; Picão *et al.*, 2010).

As AmpC são tipicamente encontradas em cromossomos de várias bactérias Gram-negativas, incluindo *Acinetobacter spp.*, *Aeromonas spp.*, *Chromobacterium violaceum*, *C. Freundii*, *Enterobacter spp.*, *E. coli*, *Hafnia alvei*, *Lysobacter lactamgenus*, *Morganella morganii*, *Ochrobactrum anthropi*, *Proteus rettgeri*, *Providencia stuartii*, *P. Aeruginosa*, *Psychrobacter immobilis*, *Rhodobacter sphaeroides*, *S. marcescens* e *Yersinia enterocolitica*. Entretanto já existem AmpC sendo detectadas em plasmídeos que geralmente codificam ESBLs (Philippon *et al.*, 2002) e frequentemente carregam outros genes de resistência como aminoglicosídeos, cloranfenicol, sulfonamida, tetraciclina, trimetoprim ou até de mercúrio (Papanicolau *et al.*, 1990; Bauernfeind *et al.*, 1996a, 1996b; Bradford *et al.*, 1997; Stapleton *et al.*, 1999). As AmpC mediadas por plasmídeos representam uma nova ameaça à saúde pois conferem resistência à 7- $\alpha$ -methoxy-cephalosporinas como cefoxitina ou cefotetan e não são afetadas por inibidores de  $\beta$ -lactamases comerciais

disponíveis. Além disso, cepas com perda de porinas na membrana externa, podem apresentar resistência às carbapenemas (Philippon *et al.*, 2002).

Em relação à classe C, bactérias do filo Proteobacteria apresentaram o maior número de  $\beta$ -lactamases desta classe. Porém, a filogenia relacionada às prováveis  $\beta$ -lactamases da classe C revelou uma diversidade destas enzimas como as subclasses AmpC, MOX-4 e CMY (Figura 5.13), sendo a subclasse AmpC a mais frequente. Esses resultados estão de acordo com os obtidos na análise por similaridade com regiões conservadas (RPS-BLAST) em bases de PSSM de genes de resistência do ARDB, no qual a subclasse ACC foi a mais abundante (53% para ZS e 20% para ZN), em ambos os hospitais (Figura 5.8). ACC (“Ambler Class C”) é um novo tipo de AmpC, pois estas eram caracterizadas por serem localizadas no cromossomo e o tipo ACC-1 (“Ambler Class C 1”) foi o primeiro tipo de AmpC descrito encontrado em plasmídeo, em *K. pneumoniae*, na Alemanha, e compartilha uma baixa identidade (> 50%) com as AmpC de patógenos avaliados (Bauernfeind *et al.*, 1999). Foi demonstrada posteriormente que a ACC-1 derivou da AmpC cromossomal de *Hafnia alvei* (Nadjar *et al.*, 2000). Até 2006 pelo menos, a ACC-1 tinha sido isolada apenas na Alemanha (Bauernfeind *et al.*, 1999), França (Girlich *et al.*, 2000; Nadjar *et al.*, 2000; Bidet *et al.*, 2005; Ohana *et al.*, 2005), Espanha (Miró *et al.*, 2005) e Tunísia (Makanera *et al.*, 2003). É preocupante saber que essa subclasse se encontra em maior abundância em ambos os hospitais, pois durante a última década têm aumentado incrivelmente sua detecção em isolados clínicos de todo o mundo, e tem sido associados ao insucesso do tratamento (Pai *et al.*, 2004; Song *et al.*, 2005) uma vez que essas enzimas não são inibidas pelo ácido clavulânico. Nas amostras de ZN a subclasse FOX também foi abundante (21%) (Figura 5.8). Esta subclasse tem sido descrita em plasmídeos (Bauernfeind *et al.*, 1997; Marchese *et al.*, 1998; Bush *et al.*, 1995), sendo a maioria susceptível às cefamicinas, como por exemplo, a cefotaxina (Bauernfeind *et al.*, 1997). Porém, já foram descritas subclasses novas resistentes como FOX-2 (em *E. coli*) (Bauernfeind *et al.*, 1997), CMY-2 (em *K. pneumoniae*) (Bauernfeind *et al.*, 1996). Como os plasmídeos que carregam os genes das subclasses da classe C geralmente também carregam genes ESBL (maioria da classe A), é preocupante saber que os genes que codificam essas enzimas são frequentes em um hospital do Rio de Janeiro, assim como a subclasse ACC.



As  $\beta$ -lactamases classe D possuem uma alta atividade hidrolítica contra cloxacilina, oxacilina e meticilina, e mutações em aminoácidos nessas enzimas também conferem resistência espectro estendido (ESBLs) e baixa susceptibilidade ao inibidor ácido clavulânico. A maioria das ESBLs classe D detectadas até hoje foram encontradas em *Pseudomonas aeruginosa* (Mugnier *et al.*, 1998). No presente trabalho, a menor quantidade de prováveis homólogos de  $\beta$ -lactamases classe D recuperadas pelo “hmmsearch” que tiveram sua função corroborada pela análise de similaridade de domínios conservados (RPS-BLAST) contra a base COG pode ser explicada pelo fato dessas enzimas possuírem uma baixa similaridade entre elas, com pouco mais de 20% de identidade entre os membros dessa subclasse, uma vez que essas foram classificadas por apresentarem a mesma atividade, em oxacilina, e substratos relacionados, não pela estrutura primária (Jacoby, 2006). De acordo com os resultados apresentados na Tabela 5.6 e Figura 5.6, pode-se perceber que a abundância de  $\beta$ -lactamases classe D foi maior nas amostras ZN, mesmo sendo menos recuperada pela metodologia descrita. A análise filogenética da classe D revelou prováveis  $\beta$ -lactamases dos esgotos hospitalares que agruparam com várias  $\beta$ -lactamases classe D produzidas por *Roseburia* que são comensais do intestino humano e pertencentes ao filo Firmicutes (Figura 5.14). As subclasses mais encontradas foram as OXA-10, OXA-2 e LCR-1 (em ambas as amostras de esgotos), de acordo com a classificação do ARDB. As OXAs compõem 5 grupos, de I ao V. O grupo I inclui a OXA 5, 7, 10 e 13 e seus derivados de espectro estendido (OXA-11, 14, 16, 17, 19). O grupo II inclui a OXA-2, 3, 15 e 20; o grupo III inclui a OXA-3, 4, 30 e 31, o grupo IV inclui apenas a OXA-9, enquanto o grupo V inclui apenas a LCR-1 (Bert *et al.*, 2002). Felizmente nenhuma dessas três enzimas mais frequentes nas amostras hospitalares são ESBLs. No entanto, já foram descritas diversas variantes de OXA-10 (OXA-11, OXA-14, OXA-16, OXA-17 e OXA-19) e OXA-2 (OXA-15) que possuem espectro estendido aos seus substratos (ESBLs) (Hall *et al.*, 1993; Danel *et al.*, 1995, 1997, 1999; Mugnier *et al.*, 1998; Poirel *et al.*, 2001). Para a comprovação de que essas prováveis OXAs são as variantes OXA-2 e OXA-10 e não uma variante derivada destas, que seriam ESBLs, experimentos laboratoriais devem ser realizados.

Em todas as três filogenias das classes A, C e D, os projetos que prevaleceram foram obtidos da base IMG/M. Tal acontecimento pode ser

explicado devido à montagem dessas sequências em “contigs”, podendo levar a um aumento do número de sequências inteiras de  $\beta$ -lactamases na busca com o “hmmsearch”, utilizando os perfis dos motivos específicos de cada classe. Em relação às prováveis  $\beta$ -lactamases dos projetos relacionados à comunidade microbiana intestinal de artrópodes que possuem interação simbiótica com fungos, todas as sequências representativas pertenciam ao filo Proteobacteria (sendo maioria da classe Gammaproteobacteria), incluindo as prováveis  $\beta$ -lactamases de estação de tratamento de esgoto (*K. pneumoniae* no dendrograma da classe C e espécies de *Burkholderia* no dendrograma da classe D). Esses resultados indicam que a diversidade taxonômica desses ambientes difere bastante dos esgotos hospitalares que contêm maioria de Bacteroidetes e Firmicutes abundantes em fezes humanas.

Baseado nas análises filogenéticas realizadas para o estudo das três classes de  $\beta$ -lactamases, dois padrões puderam ser observados, de uma forma geral. O primeiro inclui as sequências metagenômicas filogeneticamente relacionadas com as  $\beta$ -lactamases de referência (as curadas), que contribui no entendimento de sua origem e classificação, e o segundo inclui as sequências metagenômicas que não se agruparam próximas com nenhuma  $\beta$ -lactamase utilizada como referência, sugerindo um grande reservatório dessas prováveis enzimas para serem exploradas nas bases de metagenoma disponíveis.

Embora sejam utilizados na terapêutica humana, os antibióticos são extensivamente utilizados na produção animal, de forma profilática e terapêutica, ou mesmo como promotores de crescimento (aditivo na ração). Esses resíduos oriundos tanto de ambientes clínicos como de fazendas e plantações podem conter antibióticos e GRAs que eventualmente irão contaminar os ambientes naturais, levando à seleção de bactérias resistentes. Dessa forma, os mesmos genes encontrados em ambientes clínicos acabam sendo disseminados entre os ecossistemas intactos, sem nenhum registro de contaminação por antibióticos (Martinez, 2009).

A partir dos resultados encontrados no presente trabalho verificou-se que os pHMM foram sensíveis e eficazes na busca por homólogos distantes, mesmo utilizando basicamente sequências de  $\beta$ -lactamases do filo Proteobacteria (principalmente Gammaproteobacteria) para a construção dos perfis de cada classe (A, C e D). Sendo assim, a metodologia utilizada nesta etapa do trabalho foi capaz de minerar prováveis  $\beta$ -lactamases das bases de

metagenomas ambientais (CAMERA, IMG/M) e dos esgotos hospitalares, possibilitando o estudo da diversidade destas enzimas em diferentes ambientes e contribuindo com novos dados gerados no estudo do resistoma.

### **6.5 Potencial metabólico da comunidade bacteriana presente nas amostras de esgotos dos hospitais da Zona Norte e Zona Sul**

Os resultados obtidos a partir do item 4.13.1 foram anotados usando a classificação funcional da base SEED (Overbeek *et al.*, 2005), através do programa MEGAN 4. SEED é uma base curada de genomas microbianos que possui uma ferramenta de anotação funcional através de uma abordagem de subsistemas. Um subsistema é um conjunto de papéis funcionais ou grupos de genes que funcionam juntos como os genes cujos produtos estão envolvidos em uma via metabólica, ou um grupo de genes cujos produtos formam uma estrutura celular. Assim com a glicólise possui um conjunto de papéis funcionais como as enzimas que compõem essa via metabólica (glicoquinase, glicose-6-fosfato isomerase, fosfofrutoquinase etc), ou seja, possui um conjunto de vias metabólicas. A minoria das proteínas foi atribuída em um subsistema pela base SEED, uma vez que ainda existe relativamente pouca informação disponível nas bases de dados curadas. O perfil metabólico da comunidade microbiana dos esgotos hospitalares deste trabalho apresentou metabolismo de carboidratos com uma maior porcentagem. Esse resultado é semelhante a outros trabalhos, mas de estações de tratamento (Sanapareddy *et al.*, 2009). A predominância de enzimas ligadas ao metabolismo de carboidratos é esperada, já que a maioria das bactérias comensais do intestino humano é fermentadora e o intestino humano é rico em genes envolvidos no metabolismo de amido e sacarose, além do metabolismo de glicose, galactose, frutose, arabinose, manose e xilose. As bactérias pertencentes ao filo Bacteroidetes estão relacionadas com a degradação de biopolímeros, em particular os polissacarídeos. A fermentação dos polissacarídeos libera cadeias pequenas de ácidos graxos, voláteis, principalmente acetato, propionato e butirato que são absorvidos pelo hospedeiro (Thomas *et al.*, 2011).

A abundância do subsistema de virulência ficou em segundo lugar praticamente, junto com o subsistema de proteínas (Figura 5.34). Relacionado a esse subsistema o mecanismo de captação de ferro apareceu em primeiro

lugar. As bactérias precisam de ferro para seu crescimento. O ferro é um dos elementos mais abundantes da Terra, mas é pouco disponível para as bactérias em condições anaeróbicas (solubilidade limite  $10^{-17}$ M para o Fe(III) em pH ~7) e/ou em hospedeiros de mamíferos ( $10^{-26}$ M) (Otto *et al.*, 1991). Logo, para obter o ferro as bactérias desenvolveram diferentes mecanismos, sendo o mais comum os sideróforos que são quelantes de ferro (Palyada *et al.*, 2004). Bactérias patogênicas e comensais possuem sistemas de captação de ferro de alta afinidade que permitem a obtenção de quantidades suficientes desse metal essencial em um ambiente aonde o ferro é limitante (hospedeiro vertebrado) (Bullen, 1981; Crosa, 1984; Griffiths, 1987). Neste trabalho, os mecanismos mais abundantes de captação de ferro foram: aquisição de ferro por *Vibrio*, metabolismo de ferro de *Campylobacter* e mecanismos de transporte de ferro (dado não mostrado), em ambas as amostras hospitalares. Os mecanismos de captação de ferro em *Vibrio* incluem o transporte de heme dependente de Ton-B, sideróforos endógenos como “vibriobactin” e outros sideróforos produzidos por outros micro-organismos. Em relação a esse subsistema, as proteínas mais abundantes foram aquelas relacionadas com o receptor dependente Ton-B e proteína B de transporte de ferro ferroso, que também foi a mais abundante no subsistema de mecanismo de captação de ferro de *Campylobacter*. Em bactérias Gram-negativas o sistema de captação de ferro geralmente consiste em transportadores de membrana externa dependentes de Ton-B, entre outros (proteínas que se ligam a membrana periplasmática e um complexo ABC de permease de membrana interna) (Cartron *et al.*, 2006). Ton-B é uma importante proteína de translocação de energia para sistemas de captação de ferro.

É interessante destacar que foram identificados muitos genes associados à virulência, incluindo de captação de ferro em *Campylobacter* e *Vibrio*, porém estes micro-organismos, apesar de terem sido detectados nas amostras, não foram abundantes (~5% e 10% de abundância relativa do grupo Proteobacteria, para ZS e ZN respectivamente). Uma possibilidade seria que esses dois mecanismos de captação de ferro se assemelham muito com os de bactérias pertencentes aos grupos Bacteroidetes e Firmicutes como *Bacteroides* e *Eubacterium*. Porém, mais estudos devem ser realizados (baseados em análise de similaridade e filogenia) para verificar a similaridade entre os genes responsáveis por esses mecanismos de forma a elucidar tal hipótese.

Os sistemas de transporte Ton e Tol são acoplados a uma força próton motiva através da membrana interna, agindo como transdutores de energia para o transporte ativo (Ton) ou manutenção da integridade do envelope externo da bactéria (Tol) (Llobès *et al.*, 2001). Esse sistema é importante para a obtenção do ferro que está ligado à virulência da bactéria. O terceiro grupo mais abundante do subsistema de virulência é composto por proteínas relacionadas à regulação desta. Nesse subsistema apareceram apenas sequências relacionadas ao metabolismo do “Regulon Mga” de *Streptococcus*. Mga é um regulador de virulência ubíquo que controla genes (Mga regulon) cujos produtos são necessários para a aderência, internalização e evasão do sistema imune do hospedeiro. Sendo assim, esses genes encontrados em abundância podem não estar apenas relacionados ao sistema de virulência de espécies de *Streptococcus*, mas também com outras bactérias patogênicas Gram-positivas, ou até mesmo com os comensais, já que existem poucos estudos relacionados a outros micro-organismos.

## 6.6 Considerações finais

Apesar do grande problema de resistência aos antibióticos em bactérias patogênicas presente em ambientes clínicos, pouco se sabe sobre a diversidade, dispersão e origem dos genes de resistência, especialmente envolvendo a maioria das bactérias ambientais não cultiváveis. Sabe-se que o despejo de esgotos nos recursos aquáticos também tem contribuído para a disseminação dos mecanismos de resistência, e vários países descrevem a presença de bactérias resistentes em rios e esgotos (Dhanji *et al.*, 2010; Fuentes *et al.*, 2010; Ferreira *et al.*, 2011; Fontes *et al.*, 2011), demonstrando o risco do estabelecimento de patógenos resistentes no meio ambiente. A mobilização dos genes de resistência por meio de elementos genéticos, isto é, plasmídeos, transposons e integron está intimamente relacionada com a ampla disseminação das  $\beta$ -lactamases (Weldhagen, 2004), facilitando a troca de elementos genéticos entre as diversas espécies bacterianas que, posteriormente, poderão colonizar diferentes hospedeiros e ecossistemas e se disseminar por diferentes vias como esgotos, plantações, alimentos, água e animais (Silva & Lincopan, 2012). O estudo do reservatório dos genes de resistência a antibióticos através da metagenômica contribui no fornecimento de

informações sobre possíveis futuros mecanismos de resistência em ambientes clínicos. Os dados de sequências metagenômicas podem ser utilizados para gerar perfis de outras enzimas que conferem resistência a antibióticos, ajudando no entendimento da ecologia da comunidade microbiana, assim como na dispersão dos genes de resistência entre diferentes ambientes, contribuindo para o avanço e maior compreensão dos mecanismos de resistências dessas enzimas.

A partir dos resultados obtidos nesse trabalho é necessário realizar testes experimentais que confirmem as funções dessas enzimas encontradas, assim como aquelas subclasses que parecem ser ESBLs. De forma a complementar esse estudo, é interessante também clonar esses genes presentes no metagenoma dos esgotos hospitalares em diferentes vetores de expressão com o intuito de confirmar a existência de novos genes de resistência à  $\beta$ -lactâmicos que a metodologia “In silico” foi capaz de encontrar.

## **6.7 Perspectivas futuras**

Muitos genes de resistência descobertos através das abordagens independentes de cultivo parecem ser novos e distantes evolutivamente aos genes de resistência de isolados patogênicos próximos filogeneticamente. Sendo assim, uma possibilidade de estudo é aproveitar os dados de sequenciamento dos esgotos do Rio para “buscar”, por similaridade, os diferentes elementos genéticos móveis (EGMs) associados com os genes de resistência à  $\beta$ -lactâmicos como plasmídeos, transpósons e integrons com aqueles existentes nas bases de dados de EGMs, de forma a verificar sua proximidade com os EGMs de bactérias patogênicas humanas. É necessário também realizar testes experimentais para confirmar a função dessas enzimas mineiradas e verificar se essas prováveis  $\beta$ -lactamases realmente são ESBLs.

Ainda, este trabalho possibilitou a abertura para diferentes perguntas e questionamentos nessa linha de pesquisa permitindo, assim, sua continuidade. Uma possibilidade é verificar a diversidade das metalo  $\beta$ -lactamases presentes nas amostras hospitalares obtidas e a confirmação da função dessas prováveis serina  $\beta$ -lactamases mineradas e as futuras metalo-  $\beta$ -lactamases encontradas. A função dessas enzimas pode ser realizada através do desenho de iniciadores a partir das sequências de DNA metagenômico obtidas das amostras de

esgotos, com provável função de  $\beta$ -lactamases, amplificação das diferentes classes (A, C e D), clonagem e expressão. A partir da obtenção dos clones, a função e a confirmação de sua classificação pode ser confirmada por métodos bioquímicos e posterior sequenciamento dos clones, utilizando análises de bioinformática (alinhamento múltiplo para verificar as mutações pontuais, filogenia, modelagem de proteínas). De forma a complementar esse estudo, é interessante também clonar esses genes presentes no metagenoma dos esgotos hospitalares em diferentes vetores de expressão (como em *Streptomyces*) com o intuito de analisar a existência de novos genes de resistência à  $\beta$ -lactâmicos que a metodologia “In silico” não foi capaz de buscar.

É interessante também dar continuidade na anotação dos genes presentes nos metagenomas hospitalares do Rio, realizando a montagem em “contigs”, a identificação de ORFs e anotação destes. Isso permite ampliar o estudo e verificar diferentes genes de resistência para poder realizar um estudo geral do resistoma de esgotos hospitalares, de forma a caracterizá-lo melhor, contribuindo assim para o avanço no estudo evolutivo destas enzimas.

## 7 CONCLUSÕES

- A partir da metodologia utilizada neste trabalho, utilizando os pHMMs específicos para cada classe de serina  $\beta$ -lactamase, se mostrou eficiente na mineração de possíveis homólogos em bases de dados metagenômicos.
- As reconstruções filogenéticas permitiram classificar prováveis  $\beta$ -lactamases mineradas de projetos de metagenomas ambientais assim como de esgotos hospitalares do Rio, revelando uma grande diversidade dessas enzimas em diferentes ambientes, assim como sua dispersão entre os diferentes ambientes.
- As filogenias das classes de serina  $\beta$ -lactamases sugerem que as  $\beta$ -lactamases encontradas nos esgotos hospitalares do Rio são distantes evolutivamente das encontradas nos diferentes ambientes, principalmente aquelas produzidas por bactérias do filo Proteobacteria.
- Os resultados taxonômicos e funcionais entre as amostras de esgoto hospitalar, ZS e ZN, foram similares, indicando que, mesmo cada hospital recebendo diferentes comunidades com diferentes enfermidades, a localização (entre esses dois hospitais) e as diferentes atividades exercidas neles não pareceu influenciar muito na composição da comunidade microbiana como um todo.
- Os grupos Firmicutes e Bacteroidetes foram os mais abundantes da comunidade microbiana do metagenoma de esgotos hospitalares do Rio de Janeiro, semelhante à composição da comunidade microbiana do intestino humano, sugerindo a grande presença do material genético desses micro-organismos nesses efluentes hospitalares.
- A diferença mais acentuada entre a comunidade microbiana dos dois hospitais do Rio (ZS e ZN) foi em relação à diversidade de Proteobacteria (Gamma e Delta/Epsilon-Proteobacteria).
- Nos esgotos hospitalares analisados dos dois hospitais público do Rio verificou-se a presença de diversas  $\beta$ -lactamases pertencentes às três classes de serina  $\beta$ -lactamases (A, C e D), incluindo as ESBLs e as carbapenemases, sugerindo sua circulação nesses dois hospitais.



- ESBLs não tão comuns como PER, VEB e GES foram encontradas em ambas as amostras de esgotos hospitalares do Rio. A subclasse KPC também foi encontrada em abundância de 8% no hospital da Zona Norte (ZN), gerando uma preocupação em relação a uma possível dispersão dessas enzimas na cidade do Rio.
- A maior frequência de subclasses de  $\beta$ -lactamases como ACC, CfxA e LCR e OXA-10 nas amostras hospitalares de ZS em relação a uma heterogeneidade na distribuição das subclasses nas amostras de ZN sugere uma maior pressão seletiva nos esgotos do hospital da Zona Sul, provavelmente por parte de um maior consumo de antibióticos nesse hospital, conforme destacado na caracterização funcional.

## 8 REFERÊNCIA BIBLIOGRÁFICA

- Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. Call of the wild: antibiotic resistance genes in natural environments. *Nature Reviews Microbiology*. 2010; 8(4): 251-9.
- Allen HK, Moe LA, Rodbumrer J, Gaarder A, Handelsman J. Functional metagenomics reveals diverse beta-lactamases in a remote Alaskan soil. *ISME Journal*. 2009; 3(2):243-51.
- Altschul SF & Gish W. Local alignment statistics. *Methods in Enzymology*. 1996; 266: 460–480.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*. 1997; 25 (17):3389-402.
- Ambler RP. The Structure of  $\beta$ -lactamases. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 1980; 289: 321-331.
- Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C et al. The Marine Viromes of Four Oceanic Regions. *PLoS Biology*. 2006; 4(11): e368.
- APHA-AWWA-WPCF. *Standard Methods for Examination of Water and Wastewater*. 20Th Edition. Washington, DC: New York, 2001.
- Babic M, Hujer, AM, Bonomo, RA. What's new in antibiotic resistance? Focus on beta-lactamases. *Drug Resistance Updates*. 2006; 142-156.
- Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JL: Host-bacterial mutualism in the human intestine. *Science*. 2005; 307:1915-1920.
- Baldi P, Chauvin Y, Hunkapillar T, McClure M. Hidden Markov models of biological primary sequence information. *Proceedings of the National Academy of Science of the USA*. 1994; 91: 1059-1063.
- Baquero F, Martínez JL, Cantón R. Antibiotics and antibiotic resistance in water environments. *Current Opinion in Biotechnology*. 2008; 19(3):260-5.
- Bartlett JG. *Pocket Book of Infectious Disease Therapy*. Lippincott Williams and Wilkins, Baltimore, 2003.
- Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S et al. The Pfam protein families database. *Nucleic Acids Research*. 2004; 32: D138–D141.
- Bauernfeind A, Stemplinger I, Jungwirth R, Giamarellou H. Characterization of the plasmidic beta-lactamase CMY-2, which is responsible for cephamycin resistance. *Antimicrobial Agents and Chemotherapy*. 1996a; 40 (1): 221–4.
- Bauernfeind A, Stemplinger I, Jungwirth R, Wilhelm R, Chong Y. Comparative characterization of the cephamycinase *bla*CMY-1 gene and its relationship with other  $\beta$ -lactamase genes. *Antimicrobial Agents and Chemotherapy*. 1996b; 40:1926–1930.
- Bauernfeind A, Wagner S, Jungwirth R, Schneider I, Meyer D. A novel class C  $\beta$ -lactamase (FOX-2) in *Escherichia coli* conferring resistance to cephamycins. *Antimicrobial Agents and Chemotherapy*. 1997; 41:2041–2046.
- Beifang N, Fu L, Sun S, Li W. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics*. 2010; 11:187.
- Bert F, Branger C, Lambert-Zechovsky N. Identification of PSE and OXA beta-lactamase genes in *Pseudomonas aeruginosa* using PCR- restriction

- fragment length polymorphism. *Journal of Antimicrobial Chemotherapy*. 2002; 50: 11–18.
- Bidet P, Burghoffer B, Gautier V, Brahimi N, Mariani-Kurkdjian P, El-Ghoneimi A et al. *In vivo* transfer of plasmid-encoded ACC-1 AmpC from *Klebsiella pneumoniae* to *Escherichia coli* in an infant and selection of impermeability to imipenem in *K. pneumoniae*. *Antimicrobial Agents and Chemotherapy*. 2005; 49 (8):3562–5.
- Bishop RE & Weiner JH. Complementation of growth defect in an ampC deletion mutant of *Escherichia coli*. *FEMS Microbiology Letters*. 1993; 114(3):349-54.
- Bradford PA, Urban C, Mariano N, Projan SJ, Rahal JJ, Bush K. Imipenem resistance in *Klebsiella pneumoniae* is associated with the combination of ACT-1, a plasmid-mediated AmpC  $\beta$ -lactamase, and the loss of an outer membrane protein. *Antimicrobial Agents and Chemotherapy*. 1997; 41:563–569.
- Bradford PA. Extended-Spectrum  $\beta$ -lactamases in the 21st Century: Characterization, Epidemiology, and Detection of This Important Resistance Threat. *Clinical Microbiology Reviews*. 2001; 14(4): 933-951.
- Bullen JJ. The significance of iron in infection. *Reviews of Infectious Diseases*. 1981; 3:1127-1138.
- Bulychev A, Massova I, Miyashita K, Mobashery S. Evolution of the versatile  $\beta$ -lactam hydrolase activity: from biosynthetic enzymes to drug resistance factors. *Journal of the American Chemical Society*. 1997; 119:7619–7625.
- Bush K. Characterization of  $\beta$ -lactamase. *Antimicrobial Agents and Chemotherapy*. 1989; 33(3): 259-263.
- Bush K, Jacoby GA and Medeiros AA. A Functional Classification Scheme for  $\beta$ -lactamases and Its Correlation with Molecular Structure. *Antimicrobial Agents and Chemotherapy*. 1995; 39(6):1211–1233.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAL: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009; 25: 1972-1973.
- Cartron ML, Maddocks S, Gillingham P, Craven CJ, Andrews SC. Feo-Transport of ferrous iron into bacteria. *Biometals*. 2006; 19: 143-157.
- Castanheira M, Mendes RE, Walsh TR. Emergence of the extended-spectrum beta-lactamase GES-1 in a *Pseudomonas aeruginosa* strain from Brazil: report from the SENTRY antimicrobial surveillance program. *Antimicrobial Agents and Chemotherapy*. 2004; 48(6): 2344-5.
- Cavicchioli R, DeMaere MZ, Thomas T. Metagenomic studies reveal the critical and wide-ranging ecological importance of uncultivated Archaea: the role of ammonia oxidizers. *BioEssays*. 2007; 29(1): 11-14.
- Chagas TPG, Seki LM, Oliveira JAL, Cury JC, Dávila AMR, da Silva DM, Asensi MD. Multiresistance, beta-lactamase-encoding genes and bacterial diversity in hospital wastewater in Rio de Janeiro, Brazil. *Applied Microbiology*. 2011; 111(3):572-81.
- Chao A. Estimating the Population Size for Capture-Recapture Data with Unequal Catchability. *Biometrics*. 1987; 43:783.
- Chao A, Lee S-M. Estimating the Number of Classes via Sample Coverage. *Journal of the American Statistical Association*. 1992; 87:210.
- Chou HH & Holmes MH. DNA sequence quality trimming and vector removal. *Bioinformatics*. 2001; 17: 1093-1104.

- Ciampi PL, Ojeda SA and Pinto VO. Isolation and identification of enterobacteria in lettuces (*Lactuca sativa*) cultivated in the city of Valdivia, Chile. *Revista de Microbiologia*. 1987; 18:103-108.
- Coque TM, Oliver A, Pérez-Díaz JC, Baquero F, Cantón R. Genes encoding TEM-4, SHV-2, and CTX-M-10 extended-spectrum  $\beta$ -lactamases are carried by multiple *Klebsiella pneumoniae* clones in a single hospital (Madrid, 1989 to 2000). *Antimicrobial Agents and Chemotherapy*. 2002; 46: 500–510.
- Crosa JH. The relationship of plasmid mediated iron transport and bacterial virulence. *Annual Review in Microbiology*. 1984; 38:69-89.
- Cuzon G, Naas T and Nordmann P. KPC carbapenemases: what is at stake in clinical microbiology?. *Pathologie Biologie*. 2010; 58 (1): 39–45.
- Dalhoff A & Thomson CJ. The art of fusion: from penams and cepheems to penems. *Chemotherapy (Basel)*. 2003; 49: 105–120.
- Danel F, Hall LM, Gur D, Livermore DM. OXA-15, an extended-spectrum variant of OXA-2  $\beta$ -lactamase, isolated from a *Pseudomonas aeruginosa* strain. *Antimicrobial Agents and Chemotherapy*. 1997; 41:785–790.
- Danel F, Hall LMC, Duke B, Gur D, Livermore DM. OXA-17, a further extended-spectrum variant of OXA-10  $\beta$ -lactamase, isolated from *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*. 1999; 43(6): 1362–1366.
- Danel F, Hall LMC, Gur D, Livermore DM. OXA-14, another extended-spectrum variant of OXA-10 (PSE-2)  $\beta$ -lactamase from *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*. 1995; 39:1881–1884.
- De Champs C, Poirel L, Bonnet R, Sirot D, Chanal C, Sirot J et al. Prospective survey of  $\beta$ -lactamases produced by ceftazidime-resistant *Pseudomonas aeruginosa* isolated in a French hospital in 2000. *Antimicrobial Agents and Chemotherapy*. 2002; 46, 3031– 4.
- Devos D & Valencia A. Practical limits of function prediction. *Proteins: Structure Function and Genetics*. 2000; 41:98–107.
- Dhanji H, Murphy NM, Akhigbe C, Doumith M, Hope R, Livermore DM, Woodford N. Isolation of fluoroquinolone-resistant O25b:H4-ST131 *Escherichia coli* with CTX-M-14 extended-spectrum beta-lactamase from UK river water. *Journal of Antimicrobial Chemotherapy*. 2010; 66(3): 512-6.
- Diaz-Torres ML, McNab R, Spratt DA, Villedieu A, Hunt N, Wilson M et al. Novel tetracycline resistance determinant from the oral metagenome. *Antimicrobial Agents and Chemotherapy*. 2003; 47: 1430–1432.
- Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein FC, McManus PS, Handelsman J. Metagenomic analysis of apple orchard soil reveals antibiotic resistance genes encoding predicted bifunctional proteins. *Applied Environmental Microbiology*. 2010; 76: 4396–4401.
- Dropa M. Emergence of *Klebsiella pneumoniae* carrying the novel extended-spectrum beta-lactamase gene variants blaSHV-40, blaTEM-116 and the class 1 integron-associated blaGES-7, Brazil. *Clinical Microbiology and Infection*. 2010; 16(6): 630-2.
- Dubois V, Poirel L, Marie C, Arpin C, Nordmann P, Quentin C. Molecular Characterization of a Novel Class 1 Integron Containing blaGES-1 and a Fused Product of aac(3)-Ib/aac(6"-Ib" Gene Cassettes in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy* 2002; 46(3):638-645.

- Dubus A, Normark S, Kania M, Page MGP. The Role of Tyrosine 150 in Catalysis of beta-lactam hydrolysis by AmpC  $\beta$ -lactamase from *Escherichia coli* investigated by site-directed mutagenesis. *Biochemistry*. 1994; 33(28):8577-8586.
- Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M et al. Diversity of the human intestinal microbial flora. *Science*. 2005; 308: 1635–1638.
- Eddy SR, Mitchison G, and Durbin R. Maximum discrimination Hidden Markov Models of sequence consensus. *Journal of Computational Biology*. 1995; 2, 9-23.
- Eddy SR. Profile hidden Markov models. *Bioinformatics Review*. 1998; 14(9): 755-763.
- Ellerby LM, Escobar WA, Fink AL, Mitchinson C, Wells JA. The role of lysine-234 in beta-lactamase catalysis probed by site-directed mutagenesis. *Biochemistry*. 1990; 29(24):5797–5806.
- Emmanuel E, Perrodin Y, Keck G, Blanchard J-M, Vermande P. Ecotoxicological risk assessment of hospital wastewater: a proposed framework for raw effluents discharging into urban sewer network. *Journal of Hazardous Materials*. 2005; 117(1): 1–11.
- Evaldt FR. Caracterização Físico-Química de Esgotos de Quatro Hospitais da Cidade de Porto Alegre. Porto Alegre: UFRGS, 2005. Dissertação (Mestrado em Engenharia Ambiental). Faculdade de Engenharia, Universidade Federal do Rio Grande do Sul, 2005.
- Ewing B, Hillier L, Wendl MC, Green P. Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment. *Genome Research*. 1998; 8:175–185.
- Fajardo A, Martínez-Martín N, Mercadillo M, Galán JC, Ghysels B, Matthijs, S., Cornelis, P., Wiehlmann, L., Tummler, B., Baquero, F., Martínez, J.L. The Neglected Intrinsic Resistome of Bacterial Pathogens. *PLoS ONE*. 2008; 3(2): e1619.
- Felsenstein J. PHYLIP (PHYLogeny Inference Package) version 3.6a2. Distributed by the author, Department of Genetics, University of Washington, Seattle; 1993.
- Ferreira AE, Marchetti DP, De Oliveira LM, GEUAtti CS, Fuentefria DB, Coração G. Presence of OXA-23-producing isolates of *Acinetobacter baumannii* in wastewater from hospitals in southern Brazil. *Microbial Drug Resistance*. 2011; 17(2):221-7.
- Ferreira LQ, Avelar KE, Vieira JM, de Paula GR, Colombo AP, Domingues RM, Ferreira MC. Association between the *cfxA* gene and transposon Tn4555 in *Bacteroides distasonis* strains and other *Bacteroides* species. *Current Microbiology*. 2007; 54(5):348-53.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research Web Server*. 2011; 39:W29-W37.
- Fontes LC, Neves PR, Oliveira S, Silva KC, Hachich EM, Sato MIZ et al. Isolation of *Pseudomonas aeruginosa* co-producing metallo- $\beta$ -lactamase SPM-1 and 16S rRNA methylase RmtD-1 in an urban river. *Antimicrobial Agents and Chemotherapy*. 2011; 55(6):3063-4.
- Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. The shared antibiotic resistome of soil bacteria and human pathogens. *Science*. 2012; 337: 1107-1111.

- Fuentefria DB, Ferreira AE, Coração G. Antibiotic-resistant *Pseudomonas aeruginosa* from hospital wastewater and superficial water: are they genetically related? *Journal of Environmental Management*. 2010; 92(1): 250-5.
- Fuentefria DB, Ferreira AE, Graf T, Coração G. *Pseudomonas aeruginosa*: spread of antimicrobial resistance in hospital effluent and surface water. *Revista da Sociedade Brasileira de Medicina Tropical*. 2008; 41(5): 470-473.
- Garau G, García-Sáez I, Bebrone C, Anne C, Mercuri P, Galleni M et al. Update of the Standard Numbering Scheme for Class B  $\beta$ -lactamases. *Antimicrobial Agents and Chemotherapy*. 2004; 48(7): 2347-2349.
- Garau J. Beta-lactamases: current situation and clinical importance. *Intensive Care Medicine*. 1994; 20: 95-99.
- Ghuysen JM. Serine beta-lactamases and penicillin-binding proteins. *Annual Review of Microbiology*. 1991; 45: 37-67.
- Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, Gordon JI et al. Metagenomic analysis of the human distal gut microbiome. *Science*. 2006; 312(5778):1355- 1359.
- Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR et al. Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied Environmental Microbiology*. 2002; 68: 4301-4306.
- Girlich D, Karim A, Spicq C, and Nordmann P. Plasmid-mediated cephalosporinase ACC-1 in clinical isolates of *Proteus mirabilis* and *Escherichia coli*. *European Journal of Clinical Microbiology & Infectious Diseases*. 2000; 19:893-895.
- Girlich D, Poirel L, Leelaporn A, Karim A, Tribuddharat C, Fennewald M, Nordmann P. Molecular epidemiology of the integron-located VEB-1 extended-spectrum beta-lactamase in nosocomial enterobacterial isolates in Bangkok, Thailand. *Journal of Clinical Microbiology*. 2001; 39(1): 175-82.
- Goffin C & Ghuysen JM. Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. *Microbiology and Molecular Biology Reviews* 1998; 62: 1079-1093.
- Goh EB, Yim G, Tsui W, McClure J, Surette MG, Davies J. Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. *Proceedings of the National Academy of Sciences*. 2002; 99:17025-17030.
- Goldberg SD, Lannuccilli W, Nguyen T, Ju J, Cornish VW. Identification of residues critical for catalysis in a class C beta-lactamase by combinatorial scanning mutagenesis. *Protein Science*. 2003; 12(8): 1633-1645.
- Grenberg AE, Clesceri LS, Eaton AD. *Standard Methods: for the examination of Water and Wastewater*. 21 ed.: New York. 2005.
- Griffiths E. The iron uptake systems of pathogenic bacteria, p. 69-137. In J. J. Bullen and E. Griffiths (ed.), *Iron and infection*. John Wiley & Sons Ltd., Chichester, England. 1987.
- Guilfoile, Patrick: *Antibiotic-resistant bacteria*. New York; 2007.
- Gupta V. An update on newer beta-lactamases. *Indian Journal of Medical Research*. 2007; 126(5): 417-27.
- Gusatti CS, Ferreira AE, Fuentefria DB, Corção G. Resistência a  $\beta$ -lactâmicos em *Acinetobacter* spp isolados de efluente hospitalar no sul do Brasil. *Revista da Sociedade Brasileira de Medicina Tropical* 2009; 42(2): 183-187.

- Hall BG & Barlow M. Evolution of the serine  $\beta$ -lactamases: past, present and future. *Drug Resistance Updates*. 2004; 7: 111-123.
- Hall BG & Barlow M. Structure-based phylogenies of the serine  $\beta$ -lactamases. *Journal of Molecular Evolution*. 2003;57(3): 255-60.
- Hall LMC, Livermore DM, Gur D, Akova M, Akalin HE. OXA-11, an extended-spectrum variant of OXA-10 (PSE-2)  $\beta$ -lactamase from *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*. 1993;37:1637–1644.
- Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry and Biology*. 1998; 5:R245–R249.
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY et al. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biology*. 2009; 10, R32.
- Hata M, Fujii Y, Tanaka Y, Ishikawa H, Ishii M, Neya S et al. Substrate deacylation mechanisms of serine- $\beta$ -lactamases. *Biological & Pharmaceutical Bulletin*. 2006; 29(11): 2151-9.
- Hawkey PM & Jones AM. The changing epidemiology of resistance. *Journal of Antimicrobial Chemotherapy*. 2009; 64: i3-i10.
- Henriques IS, Fonseca F, Alves A, Saavedra MJ, Correia A. Occurrence and diversity of integrons and beta-lactamase genes among ampicillin-resistant isolates from estuarine waters. *Research in Microbiology*. 2006; 157(10):938-47.
- Heuer OE, Pedersen LB, Jensen LB, Madsen M, Olsen JE. Persistence of Vancomycin-Resistant Enterococci (VRE) in Broiler Houses after the Avoparcin Ban. *Microbial Drug Resistance*. 2002; 8(4): 355-61.
- Holland HD. Geochemistry-evidence for life on Earth more than 3850-million years ago. *Science*. 1997; 275:38–39.
- Hondorp ER & McIver KS. The Mga virulence regulon: infection where the grass is greener. *Molecular Microbiology*. 2007; 66(5): 1056-1065.
- Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Research*. 2007; 17: 377-386.
- Jacob F, Joris B, Lepage S, DEUart J, Frère J-M. Role of the conserved amino acids of the 'SDN' loop (Ser<sup>130</sup>, Asp<sup>131</sup> and Asn<sup>132</sup>) in a class A beta-lactamase studied by site-directed mutagenesis. *Biochemical Journal*. 1990; 271(2): 399–406.
- Jacoby GA & Medeiros AA. More extended spectrum  $\beta$ -lactamases. *Antimicrob. Agents Chemother*. 1991; 35:1697-1704.
- Jacoby GA.  $\beta$ -Lactamase nomenclature. *Antimicrobial Agents and Chemotherapy*. 2006; 50(4): 1123–1129.
- Jamin M, Wilkin JM, Frère JM. Bacterial DD-transpeptidases and penicillin. *Essays in Biochemistry*. 1995; 29: 1–24.
- Jansson J. 2011. Harnessing Metagenomics To Study Microbial Ecology in Soils, <http://earthsciences.typepad.com/blog/2011/07/harnessing-metagenomics-to-study-microbial-ecology-in-soils.html>. Acesso em: 11/2012.
- Jernberg C, Löfmark S, Edlund C, Jansson JK. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology*. 2010; 156(11): 3216-3223.
- Jiang X, NiY, JiangY, Yuan F, Han L, Li M et al. Outbreak of infection caused by *Enterobacter cloacae* producing the novel VEB-3 beta-lactamase in China. *Journal of Clinical Microbiology*. 2005; 43:826–831.

- Jones W. High-throughput sequencing and metagenomics. *Genome Research*. 2010; 33(4): 944-952.
- Jukes T, Cantor C. 1969. *Evolution of Protein Molecules*. Academy Press.
- Karchin R & Hughey R. Weighting Hidden Markov Models for maximum discrimination. *Bioinformatics*. 1998; 14, 772-782.
- Karplus K, Karchin R, Barrett C, Tu S, Cline M, Diekhans M et al. What is the value added by human intervention in protein structure prediction? *Proteins: Structure, Function, and Bioinformatics*. 2001; 45, 86-91.
- Karplus K, Sjölander K, Barrett C, Cline M, Haussler D, Hughey, R et al. Predicting protein structure using hidden Markov models. *Proteins*. 1997; 1: 134-139.
- Katoh M, Kuma M. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 2002; 30:3059-3066.
- Kazimierczak KA, Scott KP, Kelly D, Aminov RI. Tetracycline resistome of the organic pig gut. *Applied and Environmental Microbiology*. 2009; 75: 1717–1722.
- Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McInerney JO. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evolutionary Biology*. 2006; 6:29.
- Kelly JA, Dideberg O, Charlier P, Wery J, Libert M, Moews P et al. On the origin of bacterial resistance to penicillin: comparison of a beta-lactamase and a penicillin target. *Science*. 1986; 231, 1429–1431.
- Könönen E, Eerola E, Frandsen EVG, Jalava J, Mättö J, Salmenlinna S et al. Phylogenetic characterization and proposal of a new pigmented species to the genus *Prevotella*: *Prevotella pallens* sp. nov. *International Journal of Systematics Bacteriology*. 1998; 48: 47-51.
- Krause L, Diaz NN, Bartels D, Edwards RA, Pühler A, Rohwer F et al. Finding novel genes in bacterial communities isolated from the environment. *Bioinformatics*. 2006; 22 (14): 281–289.
- Krause L, Diaz NN, Goesmann A, Kelley S, Nattkemper TW, Rohwer F et al. Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Research*. 2008; 36:2230–2239.
- Kristiansson, E., Fick, J., Janzon, A., Grabic, R., Rutgersson, C., Weijdegård, B., Söderström, H., Larsson, J. D. G.. Pyrosequencing of Antibiotic-Contaminated River Sediments Reveals High Levels of Resistance and Gene Transfer Elements. *PLoS ONE*. 2011; 6(2): e17038.
- Krogh A, Brown M, Mian IS, Sjolander K, Haussler D. Hidden markov models in computational biology: applications to protein modeling. *Journal of Molecular Biology*. 1994; 235, 1501-1531.
- Kruse H & Sorum H. Transfer of multiple drug resistance plasmids between bacteria of diverse origins in natural microenvironments. *Applied in Environmental Microbiology*. 1994; 60: 4015- 4021.
- Kümmerer K. Drugs in the environment: emission of drugs, diagnostic aids and disinfectants into wastewater by hospitals in relation to other sources—a review. *Chemosphere*. 2001; 45: 957–969.
- Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Research*. 2007; 14:169–181.



- Langsruds S, Sidhu MS, Heir E, Holck AL. Bacterial disinfectant resistance – a challenge for the food industry. *International Biodeterioration & Biodegradation*. 2003; 51: 283-290.
- Leung M, Shannon K, Frensch G. Rarity of transferable  $\beta$ -lactamase production by *Klebsiella* species. *Journal of Antimicrobial Chemotherapy*. 1997; 39:737-745.
- Li W & Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22: 1658-1659.
- Lindahl E & Elofsson A. Identification of related proteins on family, superfamily and fold level. *Journal of Molecular Biology*. 2000; 295, 613–625.
- Liu B & Pop M. ARDB- Antibiotic Resistance Genes Database. *Nucleic Acids Research*. 2008; 10:1-5.
- Livermore DM.  $\beta$ -lactamases in laboratory and clinical resistance. *Clinical Microbiology Reviews*. 1995; 8(4): 557-84.
- Lloubès R, Cascales E, Walburger A, Bouveret E, Lazdunski C, Bernadac A et al. The Tol-Pal proteins of the *Escherichia coli* cell envelope: an energized system required for outer membrane integrity? *Research in Microbiology*. 2001; 152 (6): 523–529.
- Lobkovsky E, Billings EM, Moews PC, Rahil J, Pratt RF, Knox JR. Crystallographic structure of a phosphonate derivative of the *Enterobacter cloacae* P99 cephalosporinase: mechanistic interpretation of a  $\beta$ -lactamase transition-state analog. *Biochemistry*. 1994; 33:6762–6772.
- Lobkovsky E, Moews PC, Liu H, Zhao H, Frere JM, Knox JR. Evolution of an enzyme activity: crystallographic structure at 2-Å resolution of cephalosporinase from the ampC gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proceedings of the National Academy of Sciences*. EUA. 1993; 90: 11257–11261.
- MacNeil IA, Tiong CL, Minor C, August PR, Grossman TH, Loiacono KA et al. Expression and isolation of antimicrobial small molecules from soil DNA libraries. *Journal of Molecular Microbiology and Biotechnology*. 2001; 3:301–308.
- Majiduddina FK, Materonb IC, Palzkill TG. Molecular analysis of beta-lactamase structure and function. *International Journal of Medical Microbiology*. 2002; 292(2): 127-137.
- Makanera A, Arlet G, Gautier V, Manai M. Molecular epidemiology and characterization of plasmid-encoded beta-lactamases produced by Tunisian clinical isolates of *Salmonella enterica* serotype Mbandaka resistant to broad-spectrum cephalosporins. *Journal of Clinical Microbiology*. 2003; 41:2940–2945.
- Manchester LC, Poeggeler B, Alvares FL, Ogden GB, Reiter RJ. Melatonin immunoreactivity in the photosynthetic prokaryote *Rhodospirillum rubrum*: implications for an ancient antioxidant system. *Cellular and Molecular Biology Research*. 1995; 41:391–395.
- Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. "CDD: a database of conserved domain alignments with links to domain three-dimensional structure." *Nucleic Acids Research*. 2002; 30: 281-3.
- Mariat D, Firmesse O, Levenez F, Guimarães VD, Sokol H, Doré J et al. *The Firmicutes/Bacteroidetes ratio of the human microbiota changes with age*. *BMC Microbiology*. 2009; 9: 123.

- Martín-Cuadrado A-B, Lopez-Garcia P, Alba J-C, Moreira D, Monticelli L, Strittmatter A et al. Metagenomics of the Deep Mediterranean, a Warm Bathypelagic Habitat. *PLoS ONE*. 2007; 2(9): e914.
- Martinez JL. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proceedings of the Royal Society B*. 2009; 276: 2521–2530.
- Martínez JL. The role of natural environments in the evolution of resistance traits in pathogenic bacteria. *Proceedings of the Royal Society B*. 2009; 276: 2521-2530.
- Massova I & Mobashery S. Kinship and Diversification of Bacterial Penicillin-Binding Proteins and  $\beta$ -Lactamases. *Antimicrobial Agents and Chemotherapy* 1998; 42(1): 1–17.
- Matagne A, Lamotte-Brasseur J, Frère J-M. Catalytic properties of class A  $\beta$ -lactamases: efficiency and diversity. *Biochemical Journal*. 1998; 330: 581–598.
- Matthew M, Hedges RW & Smith JT. Types of  $\beta$ -lactamase determined by plasmids in Gram-negative bacteria. *Journal of Bacteriology*. 1979; 138: 657-62.
- Maveyraud L, Massova I, Brick C, Miyashita K, Samama JP, Mobashery S. Crystal-structure of 6 $\alpha$ -(hydroxymethyl)penicillanate complexed to the TEM-1  $\beta$ -lactamase from *Escherichia coli*: evidence on the mechanism of action of a novel inhibitor designed by a computer-aided process. *Journal of the American Chemical Society*. 1996; 118: 7435–7440.
- Medeiros AA. Evolution and Dissemination of  $\beta$ -lactamases Accelerated by Generations of  $\beta$ -lactam Antibiotics. *Clinical Infectious Diseases*. 1997; 24: S19-45.
- Miró E, Mirelis B, Navarro F. *Escherichia coli* producing an ACC-1 class C  $\beta$ -lactamase isolated in Barcelona, Spain. *Antimicrobial Agents and Chemotherapy*. 2005; 49 (2): 866–867.
- Monteiro J, Henriques APC, Santos AF. Carbapenem-resistant *Klebsiella pneumoniae* outbreak: emergence of KPC-2-producing strains in Brazil. In: Abstracts of the Forty-seventh Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago, IL, 2007. Washington, DC, EUA: American Society for Microbiology; Oral presentation C2-1929, p. 141.
- Morar M & Wright GD. The Genomic Enzymology of Antibiotic Resistance. *Annual Review of Genetics*. 2010; 44: 25-51.
- Mori T, Mizuta S, Suenaga H, Miyazaki K. Metagenomic Screening for Bleomycin Resistance Genes. *Applied and Environmental Microbiology*. 2008; 74(21): 6803-6805.
- Mugnier P, Casin I, Bouthors AT, Collatz E. Novel OXA-10-derived extended-spectrum  $\beta$ -lactamases selected *In Vivo* or *In Vitro*. *Antimicrobial Agents and Chemotherapy*. 1998; 42(12): 3113–3116.
- Naas T & Nordmann P. OXA-type  $\beta$ -lactamases. *Current Pharmaceutical Design*. 1999; 5 (11): 865-879.
- Naas T, Poirel L, Karim A, Nordmann P. Molecular characterization of In50, a class 1 integron encoding the gene for the extended-spectrum beta-lactamase VEB-1 in *Pseudomonas aeruginosa*. *FEMS Microbiology Letters*. 1999; 176:411–419.
- Nadjar D, Rouveau M, Verdet C, Donay L, Herrmann J, Lagrange PH et al. Outbreak of *Klebsiella pneumoniae* producing transferable AmpC-type beta-lactamase (ACC-1) originating from *Hafnia alvei*. *FEMS Microbiology Letters*. 2000; 187(1):35-40.

- Novais Â, Comas I, Baquero F, Cantón R, Coque TM, et al. Evolutionary Trajectories of Beta-Lactamase CTX-M-1 Cluster Enzymes: Predicting Antibiotic Resistance. *PLoS Pathogens*. 2010; 6(1): e1000735.
- Oefner C, Darcy A, Daly JJ, Gubernator K, Charnas RL, Heinze I et al. Refined crystal structure of  $\beta$ -lactamase from *Citrobacter freundii* indicates a mechanism for  $\beta$ -lactam hydrolysis. *Nature (London)*. 1990; 343: 284–288.
- Ohana S, Leflon V, Ronco E, Rottman M, Guillemont D, Lortat-Jacob S et al. Spread of a *Klebsiella pneumoniae* strain producing a plasmid-mediated ACC-1 AmpC beta-lactamase in a teaching hospital admitting disabled patients. *Antimicrobial Agents and Chemotherapy*. 2005; 49(5):2095–7.
- Otto BR, Verweij WR, Sparrius M, Verweij-van Vught AM, Nord CE et al. Human immune response to an iron-repressible outer membrane protein of *Bacteroides fragilis*. *Infection and Immunity*. 1991; 59: 2999–3003.
- Overbeek R, Begley T, Butler R, Choudhuri J, Chuang H, Cohoon M et al. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*. 2005; 33(17): 5691-5702.
- Pai H, Kang CI, Byeon JH, Lee KD, Park WB, Kim HB et al. Epidemiology and clinical features of bloodstream infections caused by AmpC-type-beta-lactamase-producing *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*. 2004; 48: 3720-3728.
- Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. Development of the Human Infant Intestinal Microbiota. *PLoS Biology*. 2007; 5(7): 177.
- Palyada K, Threadgill D and Stintzi A. Iron acquisition and regulation in *Campylobacter jejuni*. *Journal of Bacteriology*. 2004; 186, 4714–4729.
- Papanicolaou GA, Medeiros AA, Jacoby GA. Novel plasmid-mediated  $\beta$ -lactamase (MIR-1) conferring resistance to oxyimino- and  $\alpha$ -methoxy  $\beta$ -lactams in clinical isolates of *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemotherapy*. 1990; 34:2200–2209.
- Paterson DL & Bonomo RA. Extended-Spectrum  $\beta$ -lactamases: a Clinical Update. *Clinical Microbiology Reviews*. 2005; 18(4): 657-686.
- Paz M, Muzio H, Genini V, Magaleno A, Rossi S, Korol J. Aguas residuales de un Centro Hospitalario de Buenos Aires, Argentina. *Revista de Higiene y Sanidad Ambiental*. 2004; 4: 83-88.
- Pearson WR. Flexible sequence similarity searching with the FASTA3 program package. *Methods in Molecular Biology*. 2000; 132, 185–219.
- Peirano G, Seki LM, Passos VL, Pinto MC, Guerra LR, Asensi MD. Carbapenem-hydrolysing beta-lactamase KPC-2 in *Klebsiella pneumoniae* isolated in Rio de Janeiro, Brazil. *Journal of Antimicrobial Chemotherapy*. 2009; 63(2): 265-8.
- Pellegrino FL, Santos KRN, Riley LW, Moreira BM. blaGES carrying *Pseudomonas aeruginosa* from a public hospital in Rio de Janeiro, Brazil. *Brazilian Journal of Infectious Diseases*. 2006; 10(4): 251-3.
- Petrosino J, Cantu C 3rd, Palzkill T. beta-Lactamases: protein evolution in real time. *Trends in Microbiology*. 1998; 6(8): 6079-83.
- Philippon A, Arlet G, Jacoby GA. Plasmid-determined AmpC-Type  $\beta$ -lactamase. *Antimicrobial Agents and Chemotherapy*. 2002; 46(1): 1–11.
- Picão RC, Santos AF, Nicoletti AG, Furtado GH. Detection of GES-5-producing *Klebsiella pneumoniae* in Brazil. *Journal of Antimicrobial Chemotherapy*. 2010; 65(4): 796-807.
- Poirel L, Girlich D, Naas T, Nordmann P. OXA-28, an extended-spectrum variant of OXA-10  $\beta$ -lactamase from *Pseudomonas aeruginosa* and its

- plasmid- and integron-located gene. *Antimicrobial Agents and Chemotherapy*. 2001; 45(2): 447-453.
- Poirel L, Rotimi VO, Mokaddas EM, Karim A, Nordmann P. VEB-1-like extended-spectrum beta-lactamases in *Pseudomonas aeruginosa*, Kuwait. *Emerging Infectious Diseases*. 2001; 7(3): 468-470.
- Prado Tatiana. Avaliação da eficiência de um sistema de tratamento de efluente hospitalar por processo anaeróbio na remoção de coliformes, *Pseudomonas aeruginosa*, *Klebsiella pneumoniae* resistentes a antibióticos e Vírus da Hepatite A. Tese de dissertação apresentada à FIOCRUZ/ENSP como parte dos requisitos necessários para a obtenção do grau de mestre em Ciências (M.Sc.), 2007. 125 p.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. 2010; 464: 59–65.
- Ribeiro LMM. Avaliação quanto a carga poluidora dos esgotos líquidos de quatro hospitais de diferentes especialidades no município de Porto Alegre. Porto Alegre: UFRGS, 2005. Dissertação (Mestrado profissionalizante em Engenharia com ênfase em Tecnologia Ambiental e Tencologias Limpas). Faculdade de Engenharia, Universidade Federal do Rio Grande do Sul, 2005.
- Riesenfeld CS, Goodman RM, Handelsman J. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*. 2004; 6(9): 981-989.
- Rodriguez-Brito B, Rohwer F, Edwards R. An application of statistics to comparative metagenomics. *BMC Bioinformatics*. 2006; 7:162.
- Rogers MB, Parker AC, Smith CJ. Cloning and characterization of the endogenous cephalosporinase gene, *cepA*, from *Bacteroides fragilis* reveals a new subgroup of Ambler class A  $\beta$ -lactamases. *Antimicrobial Agents and Chemotherapy*. 1993; 37, 2391-400.
- Rost B. Enzyme function less conserved than anticipated. *Journal of Molecular Biology*. 2002; 318, 595–608.
- Rost B. Twilight zone of protein sequence alignments. *Protein Engineering*. 1999; 12, 85–94.
- Samraoui B, Sutton B, Todd R, Artymiuk P, Waley SG, Phillips D. Tertiary structural similarity between a class A  $\beta$ -lactamase and a penicillin-sensitive D-alanyl carboxypeptidase-transpeptidase. *Nature (London)*. 1986; 320, 378–380.
- Sanapareddy N, Hamp TJ, Gonzalez LC, Hilger HA, Fodor AA, Clinton SM. Molecular Diversity of a North Carolina Wastewater Treatment Plant as Revealed by Pyrosequencing. *Applied and Environmental Microbiology*. 2009; 75(6): 1688-96.
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci EUA*. 1977; 74 (12): 5463-5467.
- Sanschagrin F, Couture F, Levesque RC. Primary structure of OXA-3 and phylogeny of oxacillin-hydrolyzing class D beta-lactamases. *Antimicrobial Agents and Chemotherapy*. 1995; 39(4): 887-93.
- Sauvage E, Kerff F, Terrak M, Ayala JA, Charlier P. The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. *FEMS Microbiology Reviews*. 2008; 32(2): 234-58.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: open-source,

- platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environmental Microbiology*. 200; 75:7537–7541.
- Schlüter A, Krause L, Szczepanowska R, Goesmann B, Pühler A. Genetic diversity and composition of a plasmid metagenome from a wastewater treatment plant. *Journal of Biotechnology*. 2008; 136: 65-76.
- Schmieder R & Edwards R. Insights into antibiotic resistance through metagenomic approaches. 2012; 7(1): 73-89.
- Schuster-Boeckler B, Schultz J, Rahmann S: HMM Logos for visualization of protein families. *BMC Bioinformatics*. 2004; 5:7.
- Shannon CE, Weaver W. *The Mathematical Theory of Communication*. Univ. Illinois Press. 1949.
- Silva KC & Lincopan N. Epidemiologia das beta-lactamases de espectro estendido no Brasil: impacto clínico e implicações para o agronegócio. *Jornal Brasileiro de Patologia e Medicina Laboratorial*. 2012; 48(2): 91-99.
- Singh R, Saxena A, Singh H. Identification of group specific motifs in Beta-lactamase family of proteins. *Journal of Biomedical Science*. 2009; 16:109.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse J. Microbial diversity in deep sea and the underexplored, rare biosphere. *Proceedings of the National Academy Sciences U. S. A.* 2006; 103(32): 12115–12120.
- Sommer MOA, Church GM, Dantas G. The human microbiome harbors a diverse reservoir of antibiotic resistance genes. *Virulence*. 2010; 1:4, 299-303.
- Sommer MOA, Dantas G, Church GM. Functional characterization of the antibiotic resistance reservoir in the human microflora. *Science*. 2009; 325: 1128–1131.
- Song W, Moland ES, Hanson ND, Lewis JS, Jorgensen JH, Thomson KS. Failure of cefepime therapy in treatment of *Klebsiella pneumoniae* bacteremia. *Journal of Clinical Microbiology*. 2005; 43(9):4891–4894.
- Srivastava PK, Desai DK, Nandi S, Lynn AM. “HMM-ModE—improved classification using profile hidden Markov models by optimising the discrimination threshold and modifying emission probabilities with negative training sequences”. *BMC Bioinformatics*. 2007; 8:104.
- Stamatakis A. RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics*. 2006; 22(21):2688-2690.
- Stapleton PD, Shannon KP, French GL. Carbapenem resistance in *Escherichia coli* associated with plasmid-determined CMY-4  $\beta$ -lactamase production and loss of an outer membrane protein. *Antimicrobial Agents and Chemotherapy*. 1999; 43:1206–1210.
- Steele HL & Streit WR. Metagenomics: advances in ecology and biotechnology. *FEMS Microbiology Letters*. 2005; 247(2): 105–11.
- Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology*. 1996; 178: 591.
- Summanem P, Baron EJ, Citron DM, Strong CA, Wexler HM, Finegold SM. *Wadsworth Anaerobic Bacteriology Manual*. Belmont: Star Publishing Co, 1993.
- Swingley WD, Alsop EB, Falenski HD, Raymond J. The 470 Megabase Metagenome of the Bison Pool (Yellowstone National Park) Alkaline Hot Spring Outflow Channel. *Astrobiology Science Conference 2010: Evolution*

- and Life: Surviving Catastrophes and Extremes on Earth and Beyond, held April 26-20, 2010 in League City, Texas. LPI Contribution No. 1538, p.5525.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution*. 2011; 28: 2731-2739.
- Tanenbaum DM, Thiagarajan M, Rusch DB, Goll J, Yooshef S, Murphy S et al. The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Standards in Genomic Sciences*. 2010; 2: 229-237.
- Tatusov RL, Koonin EV and Lipman DJA. Genomic perspective on protein families. *Science*. 1997; 278: 631–637.
- Thomas F, Hehemann ER, Czjzek M, Michel G. Environmental and gut *Bacteroidetes*: the food connection. *Frontiers in Microbiology*. 2011; 2: 93.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 1994; 22: 4673-4680.
- Tian W & Skolnick J. How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of Molecular Biology*. 2003; 333: 863–882.
- Tian W, Arakaki AK, Skolnick J. “EFICAZ: a comprehensive approach for accurate genome-scale enzyme function inference”. *Nucleic Acids Research*. 2004; 32(21):6226– 6239.
- Torres-Cortés G, Millán V, Ramírez-Saad HC, Nisa-Martínez R, Toro N, Martínez-Abarca F. Characterization of novel antibiotic resistance genes identified by functional metagenomics on soil samples. *Environmental Microbiology*. 2011; 13(4): 1101-14.
- Tribble G.D, Parker AC, Smith CJ. The *Bacteroides* mobilizable transposon Tn4555 integrates by a site-specific recombination mechanism similar to that of the gram-positive bacterial element Tn916. *Journal of Bacteriology*. 1997; 179(8): 2731-2739.
- Tribble GD, Parker AC, Smith CJ. Transposition genes of the *Bacteroides* mobilizable transposon Tn4555: role of a novel targeting gene. *Molecular Microbiology*. 1999; 34 (2):385-94.
- Tribuddharat C & Fennewald M. Integron-mediated rifampin resistance in *Pseudomonas aeruginosa*. *Antimicrobial Agents and Chemotherapy*. 1999; 43: 960–962.
- Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW et al. Comparative Metagenomics of Microbial Communities. *Science*. 2005; 308:554–557.
- Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*. 2004; 438, 37–43.
- Tzouveleki LS, Tzelepi E, Tassios PT, Legakis NJ. CTX-M-type  $\beta$ -lactamases: an emerging group of extended-spectrum enzymes. *International Journal of Antimicrobial Agents*. 2000; 14: 137– 142.
- Ursell LK, Clemente JC, Rideout JR, Gevers D, Caporaso JG, Knight R. The interpersonal and intrapersonal diversity of human-associated microbiota in key body sites. *Journal of Allergy and Clinical Immunology*. 2012; 129: 1204–1208.

- Vahaboglu H, Oztürk R, Aygün G, Coşkun F, Yaman A, Kaygusuz A et al. Widespread detection of PER-1-type extended-spectrum beta-lactamases among nosocomial *Acinetobacter* and *Pseudomonas aeruginosa* isolates in Turkey: a nationwide multicenter study. *Antimicrobial Agents and Chemotherapy*. 1997; 41(10): 2265–2269.
- Vecchia AD, Thewes MR, Harb NR, Spilki FR. Diagnóstico sobre a situação do tratamento do esgoto hospitalar no Brasil. *Revista Saúde e Ambiente*. 2009; 10 (2): 65-70.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004; 304(5667):66-74.
- Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ. Screening a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Applied and Environmental Microbiology*. 1998; 64: 3075– 3078.
- Verlicchi P, Galletti A, Petrovi P, Barcelo, D. Hospital effluents as a source of emerging pollutants: An overview of micropollutants and sustainable treatment options. *Journal of Hydrology*. 2010; 389: 416-428.
- Vieites JM, Guazzaroni M-G, Beloqui A, Golyshin PN, Ferrer M. Molecular methods to study complex microbial communities. In: Streit WR & Daniel R. *Metagenomics: Methods and Protocols, Methods in Molecular Biology*. New York: Springer Science+Business Media, LLC; 2010. p.1-4.
- Von Sperling M. Introdução à qualidade das águas e ao tratamento de esgotos. 3. Ed. Belo Horizonte. Departamento de Engenharia Sanitária e Ambiental, UFMG, 2009.
- Von Sperling M. Princípios do Tratamento Biológico de Águas Residuárias. Vol.1-Introdução à Qualidade das Águas e ao Tratamento de Esgotos, 2ed. Belo Horizonte, DESA - Departamento de Engenharia Sanitária Ambiental da UFMG –Universidade Federal de Minas Gerais, 1998.
- Walther-Rasmussen J & Høiby N. Class A carbapenemases. *Journal of Antimicrobial Chemotherapy*. 2007; 60: 470–482.
- Walther-Rasmussen J & Hoiby N. Plasmid-borne AmpC  $\beta$ -lactamases. *Canadian Journal of Microbiology*. 2002; 48:479-493.
- Wang F, Li F, Chen G, Liu W. Isolation and characterization of novel cellulase genes from uncultured microorganisms in different environmental niches. *Microbiology Research*. 2009; 164:650–657.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2: A Multiple Sequence Alignment and Analysis Workbench. 2009. *Bioinformatics*; 25 (9) 1189-1191.
- Weldhagen GF. Integrons and  $\beta$ -lactamases - a novel perspective on resistance. *International Journal of Antimicrobial Agents*. 2004; 23(6): 556-62.
- Widmanna M, Pleissa J, Oelschlaeger P. Systematic Analysis of Metallo- $\beta$ -Lactamases Using an Automated Database. *Antimicrobial Agents and Chemotherapy*. 2012; 56(7): 3481-91.
- Wilke MS, Lovering AL, Strynadka NCJ.  $\beta$ -lactam antibiotic resistance: a current structural perspective. *Current Opinion in Microbiology*. 2005; 8:525–533.
- Williams JD.  $\beta$ -lactamases and  $\beta$ -lactamase inhibitors. *International Journal of Antimicrobial Agents*. 1999; 12: 3-7.
- Wright GD. Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced Drug Delivery Reviews*. 2005; 57:1451– 1470.

- Wright GD. Antibiotic resistance in the environment: a link to the clinic? *Current Opinion in Microbiology*. 2010; 13(5): 589-94.
- Wright GD. The antibiotic resistome: the nexus of chemical and genetic diversity. *Nature reviews Microbiology*. 2007; 5(3): 175-86.
- Yang AS & Honig B. An integrated approach to the analysis and modeling of protein sequences and structures. II. On the relationship between sequence and structural similarity for proteins that are not obviously related in sequence. *Journal of Molecular Biology*. 2000; 301, 679–689.
- Yigit H, Queenan AM, Anderson GJ, Domenech-Sanchez, Biddle JW, Steward CD, Alberti S, Bush K, Tenover FC. Novel carbapenem-hydrolyzing beta-lactamase, KPC-1, from a carbapenem-resistant strain of *Klebsiella pneumoniae*. *Antimicrobial Agents and Chemoteraphy*. 2001; 45(4): 1151-61.
- Yim G, Wang HH, Davies J. Antibiotics as signalling molecules. *Philosophical Transactions of the Royal Society of Lond, Series B, Biological Sciences*. 2007; 362:1195-1200.
- Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S et al. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Applied and Environmental Microbiology*. 2004; 70: 7229–7235.
- Zhang W, Fisher JF, Mobashery S. The bifunctional enzymes of antibiotic resistance. *Current Opinion in Microbiology*. 2009; 12: 505-511.
- Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 2000; 7(1-2):203-14.



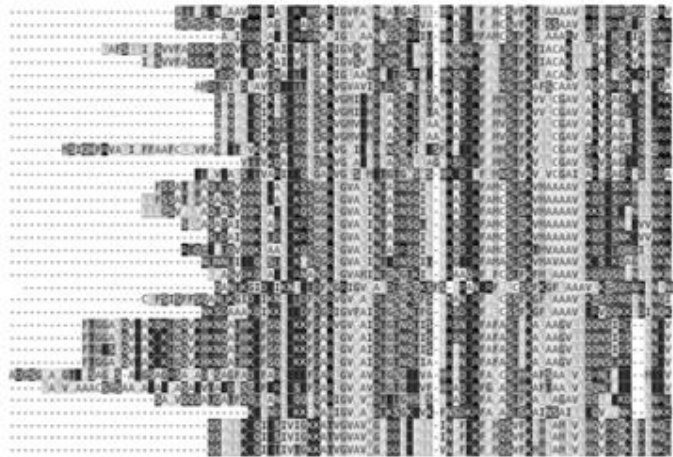
# APÊNDICE

## Alinhamentos “semente” das classes de serina $\beta$ -lactamases

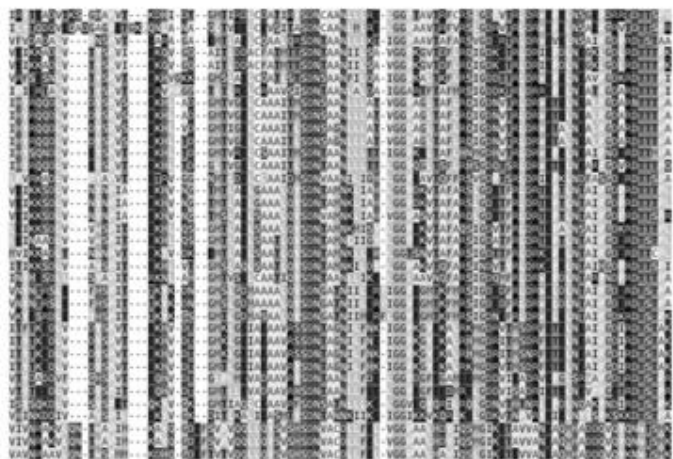
### • Classe A

### classeA.curada.uniprot.editado Certo.aln.mafft.fas

006650 Streptomyces celluloseae  
P14559 Streptomyces albus G  
Q59517 Mycobacterium fortuitum  
A4KTP8 PSE1/CARB-2 Acinetobacter nosocomialis  
Q832K4 Pasteurella multocida  
P38897 Proteus mirabilis  
Q4W1R9 Pseudomonas aeruginosa  
O08337 SMV-8 Escherichia coli  
P37323 SMV-4 Klebsiella pneumoniae  
P05192 Klebsiella pneumoniae  
A7X7N0 LEN-17 Staphylococcus epidermidis  
C8U0P5 TEM-1 Escherichia coli  
Q2VCE4 PLA-5A Raoultella planticola  
P14171 Rhodobacter capsulatus  
Q34177 CTX95 Salmonella typhimurium  
Q784F7 CTX9-2 Morganella morganii subsp.morganii  
K3K327 CTXK15 Shigella sonnei  
P22391 OXY-1 Klebsiella oxytoca  
P23954 OXY-2 Klebsiella oxytoca  
P22390 Citrobacter diversus  
P52664 Proteus vulgaris  
Q01166 Yersinia enterocolitica  
Q7ATJ4 Enterobacter cloacae  
D0V475 DMI3 Enterobacter cloacae  
P52682 Sse-1 Serratia marcescens  
Q45726 Bacillus thuringiensis  
C2RNC2 Bacillus cereus  
P90809 Bacillus cereus  
J8GXC4 Bacillus cereus  
D7AWL3 Nocardiosis dassonvillei  
Q68225 Actinomadura  
P67918 ROB-1 Haemophilus influenzae  
P08887 Staphylococcus aureus  
P37321 PER-1 Pseudomonas aeruginosa  
Q44436 PER-1 Alcaligenes faecalis  
E9L005 PER-6 Aeromonas veronii



006650 Streptomyces celluloseae  
P14559 Streptomyces albus G  
Q59517 Mycobacterium fortuitum  
A4KTP8 PSE1/CARB-2 Acinetobacter nosocomialis  
Q832K4 Pasteurella multocida  
P38897 Proteus mirabilis  
Q4W1R9 Pseudomonas aeruginosa  
O08337 SMV-8 Escherichia coli  
P37323 SMV-4 Klebsiella pneumoniae  
P05192 Klebsiella pneumoniae  
A7X7N0 LEN-17 Staphylococcus epidermidis  
C8U0P5 TEM-1 Escherichia coli  
Q2VCE4 PLA-5A Raoultella planticola  
P14171 Rhodobacter capsulatus  
Q34177 CTX95 Salmonella typhimurium  
Q784F7 CTX9-2 Morganella morganii subsp.morganii  
K3K327 CTXK15 Shigella sonnei  
P22391 OXY-1 Klebsiella oxytoca  
P23954 OXY-2 Klebsiella oxytoca  
P22390 Citrobacter diversus  
P52664 Proteus vulgaris  
Q01166 Yersinia enterocolitica  
Q7ATJ4 Enterobacter cloacae  
D0V475 DMI3 Enterobacter cloacae  
P52682 Sse-1 Serratia marcescens  
Q45726 Bacillus thuringiensis  
C2RNC2 Bacillus cereus  
P90809 Bacillus cereus  
J8GXC4 Bacillus cereus  
D7AWL3 Nocardiosis dassonvillei  
Q68225 Actinomadura  
P67918 ROB-1 Haemophilus influenzae  
P08887 Staphylococcus aureus  
P37321 PER-1 Pseudomonas aeruginosa  
Q44436 PER-1 Alcaligenes faecalis  
E9L005 PER-6 Aeromonas veronii



006650 Streptomyces celluloseae  
P14559 Streptomyces albus G  
Q59517 Mycobacterium fortuitum  
A4KTP8 PSE1/CARB-2 Acinetobacter nosocomialis  
Q832K4 Pasteurella multocida  
P38897 Proteus mirabilis  
Q4W1R9 Pseudomonas aeruginosa  
O08337 SMV-8 Escherichia coli  
P37323 SMV-4 Klebsiella pneumoniae  
P05192 Klebsiella pneumoniae  
A7X7N0 LEN-17 Staphylococcus epidermidis  
C8U0P5 TEM-1 Escherichia coli  
Q2VCE4 PLA-5A Raoultella planticola  
P14171 Rhodobacter capsulatus  
Q34177 CTX95 Salmonella typhimurium  
Q784F7 CTX9-2 Morganella morganii subsp.morganii  
K3K327 CTXK15 Shigella sonnei  
P22391 OXY-1 Klebsiella oxytoca  
P23954 OXY-2 Klebsiella oxytoca  
P22390 Citrobacter diversus  
P52664 Proteus vulgaris  
Q01166 Yersinia enterocolitica  
Q7ATJ4 Enterobacter cloacae  
D0V475 DMI3 Enterobacter cloacae  
P52682 Sse-1 Serratia marcescens  
Q45726 Bacillus thuringiensis  
C2RNC2 Bacillus cereus  
P90809 Bacillus cereus  
J8GXC4 Bacillus cereus  
D7AWL3 Nocardiosis dassonvillei  
Q68225 Actinomadura  
P67918 ROB-1 Haemophilus influenzae  
P08887 Staphylococcus aureus  
P37321 PER-1 Pseudomonas aeruginosa  
Q44436 PER-1 Alcaligenes faecalis  
E9L005 PER-6 Aeromonas veronii

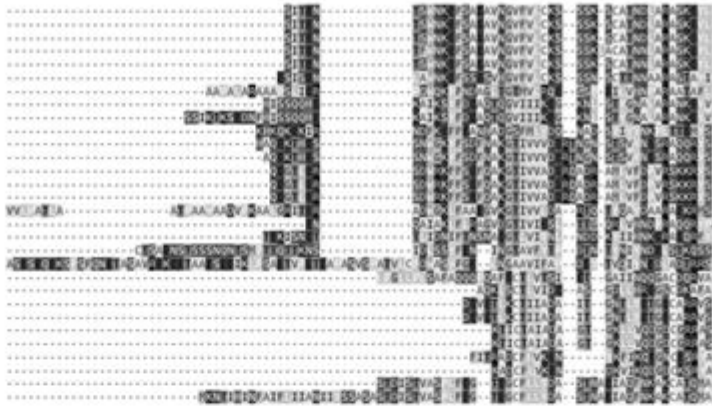




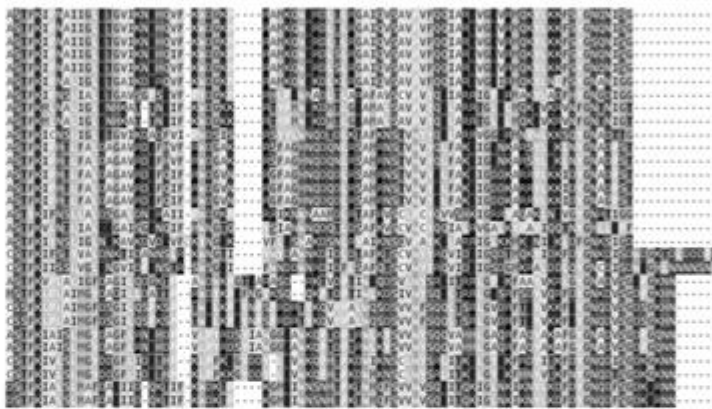
- Classe D

```
## classe0.uniprot.editada.aln.mafft.faa
```

```
P14489 OXA-10 Pseudomonas aeruginosa
B7VE58 Aeromonas hydrophila
O66778 OXA-11 Pseudomonas aeruginosa
P35695 OXA-7 Escherichia coli
O9R976 OXA-19 Pseudomonas aeruginosa
O09982 OXA-5 Pseudomonas aeruginosa
C4XK26 Desulfovibrio magneticus
O8RLA6 Acinetobacter baumannii
O2XXX9 OXA-160 Acinetobacter baumannii
Q11TW1 Cytophaga hutchinsonii
B211R4 Acinetobacter baumannii
Q7AXZ7 OXA-20 Acinetobacter baumannii
P8A1V8 OXA-2 Salmonella typhimurium
P8A1V9 OXA-2 Escherichia coli
Q51574 OXA-15 Pseudomonas aeruginosa
Q1RN99 OXA-62 Pandoraea pnomenusa
O09983 LCR-1 Pseudomonas aeruginosa
B4XT38 Brachyspira gilolicoli
D4KFP1 Meganomonas hypermegale
D6DKZ8 Clostridium saccharolyticum
Q7D988 Agrobacterium tumefaciens
O07293 OXA-18 Pseudomonas aeruginosa
P8A3M3 OXA-9 Klebsiella pneumoniae
P8A3M4 OXA-9 Enterobacter aerogenes
O6Z2U6 Burkholderia pseudomallei
O6Z2Z2 OXA-57 Burkholderia pseudomallei
Q4UK99 Rickettsia felis
Q1RKC6 Rickettsia bellii
P13661 OXA-1 Escherichia coli
Q9KHJ3 OXA-30 Shigella flexneri
```



```
P14489 OXA-10 Pseudomonas aeruginosa
B7VE58 Aeromonas hydrophila
O66778 OXA-11 Pseudomonas aeruginosa
P35695 OXA-7 Escherichia coli
O9R976 OXA-19 Pseudomonas aeruginosa
O09982 OXA-5 Pseudomonas aeruginosa
C4XK26 Desulfovibrio magneticus
O8RLA6 Acinetobacter baumannii
O2XXX9 OXA-160 Acinetobacter baumannii
Q11TW1 Cytophaga hutchinsonii
B211R4 Acinetobacter baumannii
Q7AXZ7 OXA-20 Acinetobacter baumannii
P8A1V8 OXA-2 Salmonella typhimurium
P8A1V9 OXA-2 Escherichia coli
Q51574 OXA-15 Pseudomonas aeruginosa
Q1RN99 OXA-62 Pandoraea pnomenusa
O09983 LCR-1 Pseudomonas aeruginosa
B4XT38 Brachyspira gilolicoli
D4KFP1 Meganomonas hypermegale
D6DKZ8 Clostridium saccharolyticum
Q7D988 Agrobacterium tumefaciens
O07293 OXA-18 Pseudomonas aeruginosa
P8A3M3 OXA-9 Klebsiella pneumoniae
P8A3M4 OXA-9 Enterobacter aerogenes
O6Z2U6 Burkholderia pseudomallei
O6Z2Z2 OXA-57 Burkholderia pseudomallei
Q4UK99 Rickettsia felis
Q1RKC6 Rickettsia bellii
P13661 OXA-1 Escherichia coli
Q9KHJ3 OXA-30 Shigella flexneri
```



```
P14489 OXA-10 Pseudomonas aeruginosa
B7VE58 Aeromonas hydrophila
O66778 OXA-11 Pseudomonas aeruginosa
P35695 OXA-7 Escherichia coli
O9R976 OXA-19 Pseudomonas aeruginosa
O09982 OXA-5 Pseudomonas aeruginosa
C4XK26 Desulfovibrio magneticus
O8RLA6 Acinetobacter baumannii
O2XXX9 OXA-160 Acinetobacter baumannii
Q11TW1 Cytophaga hutchinsonii
B211R4 Acinetobacter baumannii
Q7AXZ7 OXA-20 Acinetobacter baumannii
P8A1V8 OXA-2 Salmonella typhimurium
P8A1V9 OXA-2 Escherichia coli
Q51574 OXA-15 Pseudomonas aeruginosa
Q1RN99 OXA-62 Pandoraea pnomenusa
O09983 LCR-1 Pseudomonas aeruginosa
B4XT38 Brachyspira gilolicoli
D4KFP1 Meganomonas hypermegale
D6DKZ8 Clostridium saccharolyticum
Q7D988 Agrobacterium tumefaciens
O07293 OXA-18 Pseudomonas aeruginosa
P8A3M3 OXA-9 Klebsiella pneumoniae
P8A3M4 OXA-9 Enterobacter aerogenes
O6Z2U6 Burkholderia pseudomallei
O6Z2Z2 OXA-57 Burkholderia pseudomallei
Q4UK99 Rickettsia felis
Q1RKC6 Rickettsia bellii
P13661 OXA-1 Escherichia coli
Q9KHJ3 OXA-30 Shigella flexneri
```



```
P14489 OXA-10 Pseudomonas aeruginosa
B7VE58 Aeromonas hydrophila
O66778 OXA-11 Pseudomonas aeruginosa
P35695 OXA-7 Escherichia coli
O9R976 OXA-19 Pseudomonas aeruginosa
O09982 OXA-5 Pseudomonas aeruginosa
C4XK26 Desulfovibrio magneticus
O8RLA6 Acinetobacter baumannii
O2XXX9 OXA-160 Acinetobacter baumannii
Q11TW1 Cytophaga hutchinsonii
B211R4 Acinetobacter baumannii
Q7AXZ7 OXA-20 Acinetobacter baumannii
P8A1V8 OXA-2 Salmonella typhimurium
P8A1V9 OXA-2 Escherichia coli
Q51574 OXA-15 Pseudomonas aeruginosa
Q1RN99 OXA-62 Pandoraea pnomenusa
O09983 LCR-1 Pseudomonas aeruginosa
B4XT38 Brachyspira gilolicoli
D4KFP1 Meganomonas hypermegale
D6DKZ8 Clostridium saccharolyticum
Q7D988 Agrobacterium tumefaciens
O07293 OXA-18 Pseudomonas aeruginosa
P8A3M3 OXA-9 Klebsiella pneumoniae
P8A3M4 OXA-9 Enterobacter aerogenes
O6Z2U6 Burkholderia pseudomallei
O6Z2Z2 OXA-57 Burkholderia pseudomallei
Q4UK99 Rickettsia felis
Q1RKC6 Rickettsia bellii
P13661 OXA-1 Escherichia coli
Q9KHJ3 OXA-30 Shigella flexneri
```



**Número de acesso das sequências de serina  $\beta$ -lactamases utilizadas na construção dos alinhamentos “semente” para obter os pHMMs de cada classe (A, C e D), obtidas na base de proteínas não redundantes do UniprotKB, versão 09/2012.**

**Classe A:**

sp|Q06650|Streptomyces cellulosa  
sp|P14559| Streptomyces albus G  
sp|Q59517| Mycobacterium fortuitum  
tr|O34177|CTXM-5\_Salmonella typhimurium  
tr|Q704F7|CTXM-2\_Morganella morganii subsp.morganii  
tr|H3K3Z7|CTMX-15\_Shigella sonnei  
sp|P22391|OXY-1\_Klebsiella oxytoca  
sp|P23954|OXY-2\_Klebsiella oxytoca  
sp|P22390| Citrobacter diversus  
sp|P52664| Proteus vulgaris  
sp|Q01166| Yersinia enterocolitica  
tr|Q7ATJ4| Enterobacter cloacae  
tr|D0V475|IMI-3\_Enterobacter cloacae  
sp|P52682|Sme-1\_Serratia marcescens  
sp|Q45726| Bacillus thuringiensis  
tr|C2RNC2| Bacillus cereus  
sp|P00809| Bacillus cereus  
tr|J8GXC4| Bacillus cereus  
tr|D7AWL3| Nocardiosis dassonvillei  
tr|Q60225| Actinomadura  
sp|P67918|ROB-1 Haemophilus influenza  
sp|P00807| Staphylococcus aureus  
tr|A4KTP8|PSE1/CARB-2 Acinetobacter nosocomialis  
tr|Q83ZX4| Pasteurella multocida  
sp|P30897| Proteus mirabilis  
tr|Q4W1R9| Pseudomonas aeruginosa  
sp|O08337|SHV-8\_Escherichia coli  
sp|P37323|SHV-4\_Klebsiella pneumoniae  
sp|P05192| Klebsiella pneumoniae  
tr|A7X7N0|LEN-17\_Staphylococcus epidermidis  
tr|C8UQP5|TEM-1\_Escherichia coli  
tr|Q2VCE4|PLA-5A\_Raoultella planticola  
sp|P14171| Rhodobacter capsulatus  
sp|P37321|PER1\_P.aeruginosa  
tr|Q4A4J6|PER1\_Alcaldigenes faecalis  
tr|E9L0D5|PER-6\_Aeromonas veronii

**Classe C:**

tr|Q4K9D2| Pseudomonas fluorescens  
tr|Q84AE1|DHA-1\_Enterobacter cloacae

sp|P85302| *Pseudomonas fluorescens*  
tr|B9J822| *Agrobacterium radiobacter*  
tr|D0EJY7|PDC-16\_ *Pseudomonas aeruginosa*  
tr|Q8RMP1| *Pseudomonas aeruginosa\_ ampC*  
tr|C3MAK3| *Rhizobium\_ sp.*  
tr|Q7NYG4| *Chromobacterium violaceum*  
tr|I3XEC9|AmpC\_ *Sinorhizobium fredii*  
tr|Q0K1N3| *Cupriavidus necator*  
tr|A4XTY3| *Pseudomonas mendocina*  
tr|A4SK79| *Aeromonas salmonicida*  
tr|Q13NG8| *Burkholderia xenovorans*  
tr|Q53IN1|FOX-7\_ *Klebsiella pneumoniae*  
tr|Q99PX9|CMY-11 *Escherichia coli*  
tr|O52988|FOX-2 *Escherichia coli*  
tr|B2TCJ7| *Burkholderia phytofirmans*  
tr|C6B407| *Rhizobium leguminosarum*  
tr|B6VMJ3| *Photobacterium asymbiotica subsp.asymbiotica*  
tr|B5WQR5| *Burkholderia sp H160*  
tr|B2BSN6| *Aeromonas trota ampC*  
tr|A6WNL5| *Shewanella baltica*  
tr|B5ZZG7| *Rhizobium leguminosarum bv.trifolii*  
tr|Q0HIV6| *Shewanella sp.*  
tr|Q9XB24|ACC-1\_ *Klebsiella pneumoniae*  
tr|Q1I990| *Pseudomonas entomophila*  
tr|B6V743|MOX-4\_ *Aeromonas punctata*  
tr|D0SP26| *Acinetobacter junii*  
tr|B7GZL0| *Acinetobacter baumannii*  
tr|A0QZC8| *Mycobacterium smegmatis*

#### **Classe D:**

sp|P14489| OXA-10 *Pseudomonas aeruginosa*  
tr|Q7D0B8| *Agrobacterium tumefaciens*  
tr|Q8RLA6| *Acinetobacter baumannii*  
tr|Q6ZZU6| *Burkholderia pseudomallei*  
tr|Q4UK09| *Rickettsia felis*  
tr|B2I1R4| *Acinetobacter baumannii*  
tr|Q1RKC6| *Rickettsia bellii*  
tr|C4XHZ6| *Desulfovibrio magneticus*  
sp|P0A1V8| OXA-2 *Salmonella typhimurium*  
sp|P13661| OXA-1 *Escherichia coli*  
tr|Q11TW1| *Cytophaga hutchinsonii*  
sp|P35695| OXA-7 *Escherichia coli*  
sp|Q00982| OXA-5 *Pseudomonas aeruginosa*  
sp|Q06778| OXA-11 *Pseudomonas aeruginosa*  
sp|Q51574| OXA-15 *Pseudomonas aeruginosa*  
sp|P0A3M3| OXA-9 *Klebsiella pneumoniae*  
sp|P0A1V9| OXA-2 *Escherichia coli*  
tr|Q6ZZZ2| OXA-57 *Burkholderia pseudomallei*  
sp|O07293| OXA-18 *Pseudomonas aeruginosa*  
sp|P0A3M4| OXA-9 *Enterobacter aerogenes*

```
tr|D4KFP1| Megamonas hypermegale
tr|D6DKZ8| Clostridium saccharolyticum
tr|B4XT38| Brachyspira pilosicoli
tr|Q1RNF9| OXA-62 Pandoraea pnomenusa
tr|D2XKK9| OXA-160 Acinetobacter baumannii
tr|Q9KHJ3| OXA-30 Shigella flexneri
sp|Q00983| LCR-1 Pseudomonas aeruginosa
sp|Q9R976| OXA-19 Pseudomonas aeruginosa
tr|B7VES8| Aeromonas hydrophila
tr|Q7AXZ7| OXA-20 Acinetobacter baumannii
```

### **Códigos (“scripts”) utilizados:**

- Código para transformar os cabeçalhos das sequências utilizadas em números, de forma a utilizá-las como entrada para o programa RaxML-HPC (após a realização do alinhamento múltiplo).

```
#!/usr/bin/env ruby
#autor: Rodrigo Jardim

require 'rubygems'
require 'bio'

# read the sequence entry by entry through the files listed in ARGV.
sequences = Bio::FlatFile.auto(ARGV)
i=1
sequences.each do |seq|
  # do stuff with the sequence entry
  puts ">BetaLac#{i}\n#{seq.seq}\n"
  i+=1
end
```

- Código para transformar as sequências que estavam com o cabeçalho em forma de números para o cabeçalho original. Esse código é utilizado após o

processamento das sequências pelo programa RaxML-HPC, utilizando dois arquivos: o arquivo fasta original e o arquivo da árvore gerado.

```
#!/usr/bin/env ruby
#autor: Rodrigo Jardim

require 'rubygems'
require 'bio'
require 'logger'

#log = Logger.new(STDOUT)
#log.info("Initialize")
# read the sequence entry by entry through the files listed in ARGV.
#log.info("Opening tree file in #{ARGV[1]}")

fileTree = File.new(ARGV[1],"r")
arrayTree = fileTree.readlines
tree = arrayTree.to_s
fileTree.close

#log.debug("Tree: #{tree}")

#log.info("Opening sequences file in #{ARGV[0]}")
sequences = Bio::FlatFile.auto(ARGV[0])
i=1
sequences.each do |seq|
#   realName      =   seq.definition.gsub("|","_").gsub("
","_").gsub("[","_").gsub("]", "_")
   realName      =   seq.definition.gsub("|","_").gsub("
","_").gsub("[","_").gsub("]", "_").gsub(":","_").gsub("=", "_").gsub(".",
"_").gsub(";","_").gsub("(","_").gsub(")","_")
   #log.debug("Real name is #{realName} and Virtual name is
BetaLac#{i}")
   tree.gsub!("BetaLac#{i}", "#{realName},")
   tree.gsub!("BetaLac#{i}", "#{realName}")
end
```

```

        tree.gsub!("BetaLac#{i}:", "#{realName}:")
        i+=1
    end

```

```
puts tree
```

- Código para converter o arquivo do alinhamento múltiplo, em formato fasta, para formato Stockholm.

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

```
my $usage = "Usage: $0 <gapped FASTA alignment file(s)>\n";
```

```

my @argv;
while (@ARGV) {
    my $arg = shift;
    if ($arg =~ /^-/ ) {
if ($arg eq "-h") { print $usage; exit }
else { die $usage }
    } else {
push @argv, $arg;
    }
}
push @argv, "-" unless @argv;

```

```

# loop through FASTA files
foreach my $fasta (@argv) {
# read FASTA file
    my %seq;
    my @name;
    my $name;
    open FASTA, "<$fasta" or die "Couldn't open '$fasta': $!";
    while (<FASTA>) {
if (/^\s*>\s*(\S+)/) {
$name = $1;

```



```

die "Duplicate name: $name" if defined $seq{$name};
push @name, $name;
} else {
if (/^/ && !defined $name) {
warn "Ignoring: $_";
} else {
s/^s//g;
$seq{$name} .= $_;
}
}
}
close FASTA;

# check all seqs are same length
my $length;
my $lname;
foreach my $name (@name) {
my $l = length $seq{$name};
if (defined $length) {
die "Sequences not all same length ($lname is $length, $name is $l)"
unless $length == $l;
} else {
$length = length $seq{$name};
$lname = $name;
}
}

# print Stockholm output
print "# STOCKHOLM 1.0\n";
foreach my $name (@name) {
print $name, " ", $seq{$name}, "\n";
}
print "/\n";
}

```

- Código para traduzir os arquivos nos seis quadros de leitura aberta (“open reading frames”).

```
#!/bin/sh
# Autor: Rafael Cuadrat

FILE in *.fna; do transeq $FILE frame=6 `basename $FILE | sed
's/.fna/'`.traduzido; done
```

- Código para tabelar a saída de resultados do BLAST em uma planilha do excel:

```
#!/usr/bin/perl
#Autorr: Henrique Juca

use warnings;
use lib "/usr/local/bioperl-1.5.1";
use Getopt::Long;
use Spreadsheet::WriteExcel;
use Getopt::Long;
use Bio::SearchIO;
use Tie::File;

GetOptions( "table=s" => \$table_file,
           "in=s" => \$in,
           "format=s" => \$format,
           #"cog=s" => \$cog_text
           );

($table_file && $in && $format) or die
("
ERRO!: Sem arquivo(s) de Input!
-----
USAGE:
make_tabelas.pl --table=<TABLE FILE> --in=<RESULTS DIR> --
format=<INPUT FORMAT>
--cog=<LISTA DE COGS>
```

\n

#### OPTIONS:

table [obrigatorio] = nome do arquivo de saída (sem a extensão)

archive [obrigatório] = pasta onde se encontram os arquivos a serem parseados

-----

#gerando os parametros da tabela

```
my $workbook = Spreadsheet::WriteExcel->new("$table_file.xls") or die "$!";
```

```
my $sheet1 = $workbook->add_worksheet("Resultados") or die "$!";
```

```
my $geral = $workbook->add_format();
```

```
$geral->set_properties(font => 'Arial Narrow', size => '11');
```

```
my $titulo = $workbook->add_format();
```

```
    $titulo->set_font('Arial Narrow');
```

```
    $titulo->set_bold();
```

```
    $titulo->set_size(10);
```

```
    $titulo->set_align('align');
```

```
my $texto = $workbook->add_format();
```

```
    $texto->set_font('Arial Narrow');
```

```
    $texto->set_size(9);
```

```
    $texto->set_align('vjustify');
```

```
$sheet1->activate();
```

```
$sheet1->set_column(0, 0, 20);
```

```
$sheet1->set_column(1, 1, 50);
```

```
$sheet1->set_column(2, 2, 35);
```

```
$sheet1->set_column(3, 3, 50);
```

```
$sheet1->set_column(4, 4, 7);
```

```
$sheet1->set_column(5, 5, 7);
```

```

$sheet1->write(0, 0, "Query", $titulo);
$sheet1->write(0, 1, "Descricao query", $titulo);
$sheet1->write(0, 2, "GI", $titulo);
$sheet1->write(0, 3, "Descricao hit", $titulo);
$sheet1->write(0, 4, "Score", $titulo);
$sheet1->write(0, 5, "Evaluate", $titulo);

```

#inicializar contadores pra escrever os resultados corretamente na tabela

```

$count_line = '0';
$count_archive = $count_line + 1;

```

# e aqui comeca a confusão

```

opendir (ARCHIVE, "$in");
@archive_files = readdir(ARCHIVE);
close ARCHIVE;

```

#para cada arquivo dentro da pasta de resultados....

```

foreach $archive (@archive_files){
    print "inicio do $archive\n";
    if ($archive =~/(\w+)\.+/){

```

```

#&get_desc($cog_group, $cog_text, $count_archive);

```

```

#$sheet1->write($count_archive, 0, $cog_group, $texto);

```

```

    my $entao = new Bio::SearchIO( -format =>"$format",
                                   -file => "$in$archive");

```

```

    while ($result = $entao->next_result) {
        while ($hit = $result->next_hit) {

```

```

while($hsp = $hit->next_hsp ) {
    $count_archive++;
    $count_line++;
    $queryname = $result->query_name();
    $subject_id = $result->query_description();
    $hit_id = $hit->name();
    $description = $hit->description();
    $score = $hsp->score();
    $value = $hsp->evaluate();

    $sheet1->write
        ($count_line, 0, "$queryname", $texto);

    $sheet1->write
        ($count_line, 1, "$subject_id", $texto);

    $sheet1->write($count_line, 2, "$hit_id", $texto);
    $sheet1->write($count_line, 3, "$description", $texto);
    $sheet1->write_string($count_line, 4, "$score", $texto);
    $sheet1->write_string($count_line, 5, "$value", $texto);
    }
}
#@file = "";
#}
print "fim do arquivo $archive\n";
}
else {
    print "o arquivo $archive nao foi processado\n";
}

```

##SUB GET\_DESC

#subrotina pra pegar a descricao dos COGS [TEMPORARIO,  
EMERGENCIAL]

```

#sub get_desc{

#   ($cog_group, $cog_text, $count_archive) = @_;
#   tie @arquivo_cog, Tie::File, $cog_text;
#   foreach $line (@arquivo_cog){
#       #- 34 aomp--yqvdrfbcefgsnuj---- CarA [EF]
#       COG0505 Carbamoylphosphate synthase small subunit
#       if($line =~ /.+(COG\d+)\s(.+)/){
#           $COG_number = $1;
#           $COG_description = $2;

#           if ("$cog_group" eq "$COG_number"){
#               $sheet1-
#>write_string($count_archive, 1, "$COG_description", $texto);

#           }
#       }
#   }

#   }

#exit;

```

## ANEXOS

### Siglas e nomes dos projetos metagenômicos das bases CAMERA e IMG/M

#### Tipos de ambientes:

- **Aquático**

**LVIA5G:** “Freshwater microbial communities from Lake Vostok at Ice accretion”

**ASA129:** “Marine sediment archaeal communities from Santa Barbara Basin CA that are methaneoxidizing sample 912 cm ANME Sed A12 912 cm”

**OSBM58:** “Marine microbial communities from Deepwater Horizon Oil Spill sample BP Oil Spill BM58”

**GSLAS:** “Saline water microbial communities from Great Salt Lake Utah sample from South Arm Stromatolite South”

**FMCWFSCBPOS1:** “Fossil microbial community from Whale Fall at Santa Cruz Basin of the Pacific Ocean Sample \*1”

**YNP233020:** “Hot spring microbial community from Yellowstone Hot Springs sample from Nymph Lake 10”

**LWMOL:** “Methylophilic community from Lake Washington sediment Methanol enrichment”.

**BISON3:** “Yellowstone Bison Hot Spring Pool 3 Metagenome”.

**BISON4:** “ Yellowstone Bison Hot Spring Pool 4 Metagenome”

**IRNMTN\_4:** “Acid Mine, Richmond Mine, Iron Mountain (EUA), site 4”

**IRNMTN\_5:** “Acid Mine Richmond Mine, Iron Mountain (EUA), site 5, temperature 38°C.

**ALOHA:** “Marine planktonic communities from Hawaii Ocean Times Series Station (HOT/ALOHA)”.

**BATS (Sargasso)** “Sargasso Sea Planktonic Microbial Community - Bermuda Time Series BATS Station”. “Metagenome isolated from the BATS station (Bermuda Time Series Station) in the Sargasso Sea”.

- **Terrestre**

**MRS1b:** “Miscanthus rhizosphere microbial communities from Kellogg Biological Station MSU sample Replicate 1”

**MRSJC2b:** “Mesophilic rice straw”

**TRSJC2b:** “Thermophilic rice straw”

**SwBSRL2:** “Switchgrass rhizosphere microbial community from Michigan US sample from Rose Lake bulk soil RL2 “

**FACENCTA:** “Soil microbial communities from sample at FACE Site 4 Nevada Test Site Crust CO2”

**DcrCPGB:** “Soil microbial communities sample from Dark Crust Colorado Plateau Green Butt”

**MBSR1b:** “Miscanthus rhizosphere microbial communities from Kellogg Biological Station MSU”

**LWSO:** “Sediment microbial communities from Lake Washington Seattle for Methane and Nitrogen Cycles original sample replicate 1”

**SMCMF:** “Soil microbial communities from Minnesota Farm”

**PRSSG:** “Feedstockadapted consortia SG only”

**PRSSGFe:** “Feedstockadapted consortia SG + Fe”

**A5c1:** “Active Layer”

**PL\_P1:** “Permafrost Layer P1”

**RicEn:** “Endophytic microbiome from Rice”

- **Interação comunidade microbiana com o hospedeiro**

**XAGC:** “*Xyleborus affinis* microbiome from Bern Switzerland sample of gallery community”

**XyIAfBA:** “*Xyleborus affinis* microbiome from Bern Switzerland sample of adult community - Ambrosia beetle adult”

**APTFG-** “Fungus garden microbial communities from *Apterostigma dentigerum*”

**ACODB:** “Fungus garden microbial communities from *Atta colombica* in Panama sample from dump bottom”

**ACODT:** “Fungus garden microbial communities from *Atta colombica* in Panama sample from dump top”

**ACOFGB:** “Fungus garden microbial communities from *Atta colombica* in Panama- sample from fungus garden bottom”



**ACOFGT:** “Fungus garden microbial communities from *Atta colombica* in Panama-sample from fungus garden top”

**AECFG:** “Fungus garden combined”

**DPOB:** “Mountain Pine Beetle microbial communities from Grand Prairie Alberta sample from Hybrid pine”

**DPOLB:** “Mountain Pine Beetle microbial communities from McBride British Columbia Canada sample from Lodgepole pine”

**FGTW:** “Fungus growing Termite worker microbial community from South Africa sample from Oerlemans Farm”

**CLOFG:** “*Cyphomyrmex longiscapus* fungus garden”

**SPBFG:** “*Dendroctonus frontalis* Fungal community”

**HDG:** “Human Distal Gut”

**HGC:** “Human Gut Community”

- **“Engenherados”**

**PLMO:** “Wastewater treatment plant plasmid pool from Canton de Vaud Switzerland sample from Preverenges Morges”

**BMHB3:** “Biofuel Metagenome”

**MA40A:** “Mixed alcohol bioreactor microbial communities from Texas AM University sample 40°C degree reactor”

“Sludge/US\_Virion”

**SRS054956:** “hypothetical protein Human Stool microbiome from visit number 1 of subject 737052003”.

**CAPI:** “Wastewater treatment Type I Accumulibacter community from EBPR Bioreactor”.

**PFMN:** Microbioma fecal de suínos – “Fecal microbiome of Swine from Yorkshire Ohio”.

- **Ar**

**AMCSIAF1:** “Air microbial communities Singapore indoor air filter 1”

**AMCSIAF2:** “Air microbial communities Singapore indoor air filter 2”

## **Programas e algoritmos utilizados:**

Pacote HMMER 3 (“hmmbuild” e “hmmsearch”)

MAFFT (versão 6.717b)

MEGA 5.1

TrimAL (versão 1.2)

Pacote MOTHUR (contêm os algoritmos ACE, Shannon e Chao1)

Jalview (versão 2.0)

RaxML-HPC (versão 7.4.2)

Modelgenerator (0.82)

MEGAN 4

RDP (“classifier”)

CD-HIT 454 (versão 4.5.4)

Lucy (versão 1.2)

CD-HIT (versão 4.5.7)

SFF extract (versão 0.28)

TRANSEQ (Pacote EMBOSS, versão 6.1.0)

RPS-BLAST (versão 2.2.21)

BLASTP (versão 2.2.21)

BLASTN (versão 2.2.21)

PHRED

Decipher (versão de outubro de 2012)

Pacote PHYLIP 3.6 (contêm o algoritmo DNADIST)

ClustalW 1.82

LogoMat-M (Sanger Institute)