



RENORBIO

Programa de Pós-graduação em Biotecnologia

**DESENVOLVIMENTO DE UMA PLATAFORMA DE BIOINFORMÁTICA
INTEGRADA APLICADA A IDENTIFICAÇÃO MOLECULAR DE
MICRORGANISMOS PATOGÊNICOS**

Felipe José de Queiroz Sarmento

João Pessoa

Setembro de 2013

RENORBIO
Universidade Federal da Paraíba
Centro de Biotecnologia
Programa de Pós-graduação em Biotecnologia

**DESENVOLVIMENTO DE UMA PLATAFORMA DE BIOINFORMÁTICA
INTEGRADA APLICADA A IDENTIFICAÇÃO MOLECULAR DE
MICRORGANISMOS PATOGÊNICOS**

Felipe José de Queiroz Sarmento

Tese de Doutorado apresentada ao Programa de Pós-graduação em Biotecnologia, da Rede Nordeste de Biotecnologia, como parte dos requisitos necessários à obtenção do título de Doutor em Biotecnologia .

Orientador: Prof. Dr. Demetrius Antônio Machado de Araújo
Coorientador: Prof. Dr. Tiago Gomes de Andrade

João Pessoa
Setembro de 2013

S246d Sarmento, Felipe José de Queiroz.
Desenvolvimento de uma plataforma de bioinformática integrada aplicada a identificação molecular de microrganismos patogênicos / Felipe José de Queiroz Sarmento.- João Pessoa, 2013.
96f. : il.
Orientador: Demetrius Antônio Machado de Araújo
Coorientador: Tiago Gomes de Andrade
Tese (Doutorado) - UFPB/CB-RENORBIO
1. Bioinformática. 2. Banco de dados - microrganismos.
3. Integração de dados biológicos. 4. Microrganismos patogênicos. 5. Prontuário eletrônico do paciente.

UFPB/BC

CDU: 575.112(043)



Coordenação Ponto Focal do Programa de Pós-Graduação em Biotecnologia
Universidade Federal da Paraíba
Centro de Biotecnologia

Cidade Universitária, Campus I. CEP:58051-900. João Pessoa, PB.

Telefone : (83) 3216.7173 E-mails: secrenorbio@cbiotec.ufpb.br, renorbiopb@lff.ufpb.br

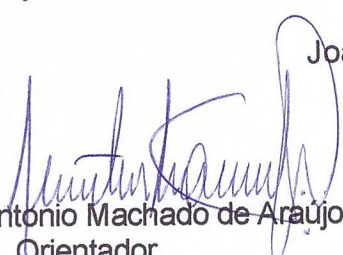
Homepage: <http://www.renorbio.org.br>

ATA DE DEFESA DE TESE DE DOUTORADO DO DOUTORANDO FELIPE JOSÉ DE QUEIRÓZ SARMENTO

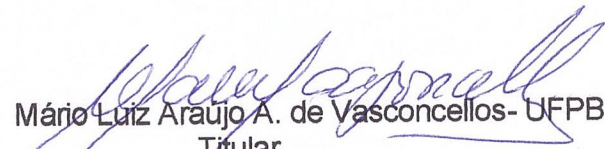
Aos vinte e sete dias do mês de fevereiro do ano de dois mil e treze, às 14 horas, reuniu-se a banca de Defesa de Tese composta pelos Professores Doutores Demétrius Antonio Machado de Araújo (orientador), Universidade Federal da Paraíba, como Presidente e os membros titulares: Mário Luiz Araújo de Almeida Vasconcellos, Universidade Federal da Paraíba, Gerd Bruno da Rocha, Universidade Federal da Paraíba, Valdir de Queiroz Balbino, Universidade Federal de Pernambuco, Ederson Akio Kido, Universidade Federal de Pernambuco, que perante a banca, o estudante regularmente matriculado no Curso de Doutorado em Biotecnologia da Rede Nordeste de Biotecnologia – RENORBIO, Ponto Focal Paraíba, defendeu, para preenchimento do requisito de doutor, sua Tese intitulada “**Desenvolvimento de uma plataforma de bioinformática integrada aplicada a identificação molecular de microrganismos patogênicos**”. Tendo o doutorando sido submetido à sabatina, dispondo cada membro da banca do tempo para tal. Finalmente, a Banca reuniu-se em separado e concluiu por considerar o doutorando **APROVADO** por sua tese e sua defesa terem, por unanimidade, emitido o conceito satisfatório.

Eu, Demétrius Antonio Machado de Araújo, que presidi a Banca de Tese, assino a presente Ata, juntamente com os demais membros e dou fé.

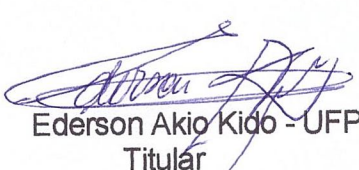
João Pessoa, 27 de fevereiro de 2013.


Demétrius Antonio Machado de Araújo- UFPB
Orientador


Gerd Bruno da Rocha- UFPB
Titular


Mário Luiz Araújo A. de Vasconcellos- UFPB
Titular


Valdir de Queiroz Balbino- UFPE
Titular


Ederson Akio Kido - UFPE
Titular

Dedicatória

Dedico esta pesquisa ao meu pai Raimundo (*in memoriam*) e a minha mãe Esmeralda.
À minha esposa, pelo apoio incondicional e as minhas filhas Anna Lara e Clarissa.

Agradecimentos

A **Deus** e a **Nossa Senhora** por me encherem de ricas bênçãos todos os dias.

Ao Prof. Dr. **Demetrius Antônio Machado de Araújo**, meu orientador, meu respeito, admiração e gratidão, e por confiar que mesmo sem liberação do trabalho que essa pesquisa fosse realizada.

Ao Prof. Dr. **Tiago Gomes de Andrade**, meu coorientador, pela confiança, paciência, amizade e orientação durante toda esta caminhada.

Aos membros da banca de qualificação, os Profs. Drs. **Fábio Correia Sampaio**, **Mário Luiz Araújo de Almeida Vasconcellos** e **Valdir de Andrade Braga**, pelas contribuições que permitiram enriquecer este trabalho;

Aos membros da banca de defesa, os Profs. Drs. **Mário Luiz Araújo de Almeida Vasconcellos**, **Gerd Bruno da Rocha**, **Valdir de Queiroz Balbino** e o **Ederson Akio Kido** pelas observações e discussões que permitiram enriquecer este trabalho, bem como as futuras versões dos sistemas desenvolvidos;

Aos meus amigos **Daniel Gomes Coimbra** e **Felipe Prata Lima**, pela amizade, estímulo, e todo o apoio no desenvolvimento dessa tese;

À técnica e pesquisadora Dra. **Tereza Cristina Grisi** e o doutorando **Itácio Padilha**, dos Laboratórios de Biologia Molecular e Ecologia (LABIMES I e II) do Departamento de Biologia Molecular da UFPB, pelos ensinamentos laboratoriais e apoio na fase de sequenciamento das amostras;

As minhas filhas, **Anna Lara** (6 anos) e **Clarissa** (11 meses), pela inocência, bondade e paciência que tiveram comigo em todos os momentos, e que no futuro entendam a minha ausência durante os momentos mais simples do dia-a-dia.

Faço as palavras da minha esposa **Patrícia Sarmiento**, as minhas, ao finalizar seu mestrado, "...pelo incentivo e confiança diários, que mesmo vivenciando seu mestrado sempre esteve ao meu lado me apoiando ...na alegria e na tristeza". Sucesso no seu doutoramento em Biotecnologia!

Ao CNPq, FAPEAL, e ao Programa de Pesquisa para o SUS (PP/SUS) pelo incentivo financeiro para a realização desta pesquisa.

A **todos** que direta ou indiretamente contribuíram para a realização deste estudo.

RESUMO

DESENVOLVIMENTO DE UMA PLATAFORMA DE BIOINFORMÁTICA INTEGRADA APLICADA A IDENTIFICAÇÃO MOLECULAR DE MICRORGANISMOS PATOGÊNICOS

A maioria das pesquisas em epidemiologia molecular, diagnóstico molecular e genética evolutiva são confrontadas com o gerenciamento de grandes volumes de dados. Além disso, os dados utilizados em estudos de doenças patogênicas são complexos e geralmente derivam de instituições tais como hospitais ou laboratórios. Embora já existam propostas que conecte informações moleculares ao diagnóstico de patogenias, nenhuma delas utilizam ferramentas de bioinformática de alto desempenho incorporadas a um sistema e vinculada a um prontuário eletrônico do paciente. MolEpi foi desenvolvido como um sistema de gerenciamento de dados e informações dimensionado a saúde pública, incorporando informações clínicas e epidemiológicas sobre pacientes e dados moleculares de sequências do gene rRNA 16S de bactérias patogênicas. Para identificação destas bactérias foram utilizadas amostras biológicas (urina, secreções e purulentas de feridas, aspirado traqueal e sangue) e PCR seguida de sequenciamento e análise da região conservada codificadora de RNA ribossômico (rDNA) 16S. Esta estratégia permite uma identificação bacteriana rápida, independente de conhecimento prévio da espécie de microrganismo em estudo. O MolEpi é um sistema facilmente atualizável com as sequências específicas de bancos como Genbank(NCBI), RDP-II (Ribosomal Database Project - MSU) e GreenGene (LBL). A partir da confirmação e validação das sequências dos isolados clínicos, estas podem ser utilizadas como referência na identificação de outros microrganismos desconhecidos. Neste sentido, foi estabelecido um banco de dados local, representativo do perfil de patógenos encontrados na unidade hospitalar de estudo e objeto de vigilância epidemiológica. Para o desenvolvimento do MolEpi, utilizamos a linguagem Java e banco de dados PostgreSQL8.3. Foi desenvolvido também o BACSearch, que possui os seguintes programas: para o processamento de sequências de rDNA 16S utilizamos os frameworks BioJava; para alinhamento múltiplo foi implementado o ClustalW2, MAFFT e o MUSCLE e para edição do alinhamento múltiplo e análise filogenética foi utilizado JalView®2.4.0b2. O sistema foi validado com 200 espécimes clínicos identificadas e isoladas de sítios de infecção hospitalar. As sequências de DNA produzidas a partir destas amostras foram submetidas ao BLAST, utilizando a ferramenta desenvolvida, identificando *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Klebsiella pneumoniae* e *Staphylococcus aureus* como os principais patógenos correspondentes. Os dados sobre o padrão de resistência das espécies foram obtidos em laboratório de microbiologia e incorporados ao banco de dados. A aplicação do MolEpi ao Sistema Único de Saúde poderá fornecer diagnósticos mais rápidos, precisos, e interligados a uma rede de informações relevantes para o profissional de saúde.

Palavras-chave: Bioinformática, Banco de dados de microrganismos, Integração de dados biológicos, Microrganismos patogênicos, Prontuário eletrônico do paciente.

ABSTRACT

DEVELOPMENT OF AN INTEGRATED BIOINFORMATICS PLATFORM APPLIED TO MOLECULAR IDENTIFICATION OF PATHOGENIC MICROORGANISMS

Various researches in molecular epidemiology, molecular diagnosis and evolutionary genetics related to pathogens are compared to managing large amounts of data derived from institutions such as, hospitals or laboratories. Although there already are some proposals to connect molecular information to the diagnosis of pathogens, none of them uses high performance bioinformatics tools which are embedded in a system and linked to a patient's electronic record. The MolEpi tool has been developed as a system of data and information management addressed to public health, incorporating clinical and epidemiological information about patients, as well as molecular data of 16S rRNA sequences of pathogenic bacteria. In order to confirm which species of these bacteria were identified, biological samples (urine, secretions and purulent wounds, tracheal aspirate and blood) and subsequently incubation and growth of colonies in culture, and PCR was used followed by sequencing and analysis of the conserved coding region for 16S ribosomal RNA (rDNA). Such strategy enabled fast bacterial identification, regardless of prior knowledge of the species of microorganism under study. Moreover MolEpi is a system interconnected to repositories of specific sequences as Genbank (NCBI), RDP-II (Ribosomal Database Project - MSU) and GreenGene (LBL). In this way, once the sequences of clinical isolates are confirmed and validated, they can be used as reference in the identification of other unknown microorganisms. Thus, a local database was established, representing the profile of pathogens found in the hospital unity of study and which should be object of public health surveillance. In order to develop MolEpi, we used the Java programming language and the PostgreSQL8.3 object-relational database. It was also developed BACSearch, which has the following programs to handle the analysis of 16S rDNA sequences, we used the framework BioJava; to multiple alignment, ClustalW2, MAFFT and MUSCLE, and for editing of multiple alignment and phylogenetic analysis, the JalView2.4.0 was used. The system was validated with 200 clinical specimens isolated and identified from sites of nosocomial infection. The DNA sequences produced from these samples were subjected to BLAST by using the developed tool, which identified *Pseudomonas aeruginosa*, *Acinetobacter baumannii*, *Klebsiella pneumoniae* and *Morganella morganii* as the main pathogens involved. Data on resistance patterns of the species were obtained in microbiology laboratory, and incorporated into the database. The application of MolEpi tool to the Health System can provide prompt and accurate diagnosis, connected to relevant network information which can be intended for health professionals.

Keywords: Bioinformatics, Database of microorganisms, Biological databases integration, Pathogenic microorganisms, Electronic patient health records.

Lista de Figuras

2.1	Cadeia de transmissão de infecção hospitalar. Modificado de SEHULSTER et al. (2003).	6
2.2	O gene rRNA 16S. A - relação entre o rRNA 30S (subunidade menor, <i>small subunit</i> (LSU))e 50S (subunidade maior, <i>large subunit</i> (LSU)), bem como as regiões intergênicas ITS. B e C - As nove regiões variáveis e o destaque para as regiões V1, V2 e V3 (modificado Yeang et al. (2007)	11
2.3	Comparação entre o alinhamento global e local (Modificado de Thompson, Plewniak e Poch (1999b)).	15
2.4	Pontuação do alinhamento local (Thompson, Plewniak e Poch (1999b)).	15
2.5	Alinhamento múltiplo de sequências de <i>P. aeruginosa</i> . Autor, 2012.	17
2.6	Crescimento dos bancos de dados DDBJ/EMBL/GenBank versão 88 (Adaptado de(KODAMA et al., 2012)).	21
2.7	Crescimento do banco de dados de rRNAs. (PRUESSE et al., 2007)	22
2.8	Bioinformática translacional união do conhecimento biológico e o clínico. Adaptado de (SARKAR et al., 2011)	24
4.1	O diagrama exemplificando a relação entre Model, View e Controller - MVC, demonstrando as associações diretas e indiretas.	30
4.2	Mapa do estado de Alagoas.	33
5.1	Interface gráfica do software BACSearch. A - interface para inserção de sequencias; B - resultado do alinhamento do BLAST; C - lista dos resultados da busca por similaridade, demonstrando <i>score</i> e o <i>e-value</i> ; D - agrupamento de resultados para alinhamento múltiplo; e E - resultado do alinhamento múltiplo de sequências para visualização dos editores de alinhamento e estudo filogenético.	39
5.2	Fluxo de trabalho do BACSearch. A - Entrada da sequência; B - Alinhamento integrado de sequência através do BLAST em bancos de dados de referência; C - Bancos de dados de referência; D - O resultado do alinhamento do BLAST são inseridos para o estudo de MSA; E - Programas utilizados no MSA; F - Os resultados do MSA são utilizados em estudos de árvores filogenéticas, com Jalview e/ou Archeopyterxy.	40
5.3	Modelo conceitual Entidade-Relacional do banco de dados MolEpi. A - Dados pessoais do paciente para iniciar o processo de análise; B - Nesse módulo do banco de dados que apresenta todas as tabelas que irão armazenar as informações das amostra obtidas e estão relacionadas com o registro eletrônico do paciente; C - Os registros dos alinhamentos obtidos são armazenados com o melhor resultado e o nome do banco de dados biológico; D - Todos os dados relacionados com a análise da amostra e que são realizados no laboratório de microbiologia são armazenados nesse conjunto de tabelas; E - A identificação realizada até o táxon mais específico obtido através dos alinhamentos são armazenados nesse conjunto de tabelas.	50
5.4	Fluxo de trabalho do MolEpi. A - Estudo das relações epidemiológicas no ambiente hospitalar; B - Informações do paciente; C - Inserção de informação das amostras biológicas; D - Estudo microbiológico da amostra em laboratório; E - Inserção do produto do sequenciamento; F - Identificação e análise do produto do sequenciamento; G - Informações centradas no Prontuário Eletrônico do Paciente; H - Recuperação de informações para análise das sequências; I - Identificação através do BLAST; H - Integração do BLAST como bancos de dados de referência, em L; M - Estudo das sequências através da interface de análise filogenética, que utilizam as informações armazenadas no registro eletrônico do paciente em N.	51

Lista de Tabelas

4.1	Espécie utilizada na validação do software BACSearch.	32
4.2	Reação da PCR para amplificação parcial do gene rRNA 16S.	35
5.1	Identificação obtida através do software BACSearch.	44
5.2	Espécies identificadas através do MolEpi.	45

Lista de Abreviaturas e Siglas

ANVISA	Agência Nacional de Vigilância Sanitária
ARS	Agricultural Research Service
ATCC	American Type Culture Collection
BHI	Brain Heart Infusion
BLAST	Basic Local Alignment Search Tool
CCIH	Comissão de Controle de Infecção Hospitalar
CDC	Center for Disease Control and Prevention
CLSI	Clinical and Laboratory Standards Institute
CTAB	Brometo de Cetiltrimetilamônio
DDBJ	DNA Data Bank of Japan
EMBL	European Molecular Biology Laboratory
ESBL	Beta-Lactamases de Espectro Ampliado
E-Value	Expect Value
IC	Infecção Comunitária
IH	Infecção Hospitalar
INPI	Instituto Nacional de Propriedade Industrial
INSDC	International Nucleotide Sequence Database Collaboration
ITU	Infecção do Trato Urinário
LABMEG	Laboratório de Biologia Molecular e Expressão Gênica
LBL	Lawrence Berkeley National Laboratory
MAS	Alinhamento Múltiplo de Sequências
MDR	Multirresistente a drogas
MLS	Macrolídeos, Lincosamidas e Estreptograminas B
MRS	<i>Staphylococcus</i> Coagulase-Negativos Resistentes à Meticilina
MRSA	<i>Staphylococcus aureus</i> Resistentes à Meticilina
MSU	Michigan State University
NAR	Nucleic Acids Research
NCBI	National Center of Biotechnology Information
NCCLS	National Committee on Clinical Laboratory Standards
NHSN	National Healthcare Safety Network
OMS	Organização Mundial de Saúde
OTU	Unidade Taxonômica Operacional
PCR	Reação em Cadeia da Polimerase
PEP	Prontuário Eletrônico do Paciente
RDP-II	Ribossomal Database Project II
rRNA	RNA ribossomal
SIH	Sistema de Informação Hospitalar do Ministério da Saúde
SUS	Sistema Único de Saúde
SVD	Sonda Vesical de Demora
UEA	Unidade de Emergência do Agreste
UFC	Unidades Formadoras de Colônias
UTI	Unidade de Terapia Intensiva

Sumário

	Página
Lista de Figuras	g
Lista de Tabelas	h
Lista de Abreviaturas e Siglas	i
1 Introdução	1
2 Fundamentação Teórica	4
2.1 Infecção hospitalar	4
2.2 Biologia de bactérias	7
2.2.1 Genômica de Bactéria	7
2.2.2 Taxonomia e evolução de bactérias	8
2.2.3 Identificação molecular de microrganismos através do rRNA e sua importância	9
2.3 Bioinformática na identificação de bactérias	13
2.3.1 Alinhamento de sequências	14
2.3.2 Alinhamento múltiplo de sequências	16
2.3.3 Bancos de dados biológicos	20
2.3.4 Bancos de dados de rRNA	21
2.4 Bioinformática aplicada a saúde	23
2.4.1 Sistema de diagnóstico em saúde	24
2.4.2 Sistemas automatizados aplicados à saúde	25
3 Objetivos	28
3.1 Gerais	28
3.2 Específicos	28
4 Material e Métodos	29
4.1 Levantamento de requisitos	29
4.1.1 Desenvolvimento dos sistemas	29
4.2 Metodologia de validação	31
4.2.1 BACSearch	31
4.2.2 MolEpi	32
5 Resultados e Discussão	37
5.1 Sistemas desenvolvidos	37
5.1.1 BACSearch	37
5.1.2 MolEpi	40
5.2 Validação do BACSearch	42
5.3 Validação do MolEpi	44
6 Conclusões	52
6.1 Considerações finais	52
6.2 Trabalhos futuros	53
Referências	55
A Apêndice - Formulário de coleta de dados do paciente	66

B	Apêndice - Termo de compromisso livre e esclarecido	69
C	Apêndice - Sequências depositadas no NCBI	72
D	Apêndice - Resultado do alinhamento contra os bancos de dados de referência	78
E	Apêndice - Interface das principais aplicações do MolEpi	84
1	Anexo - Paracer do Comitê de Ética	91
2	Anexo - Registros Instituto Nacional de Propriedade Industrial - INPI	93

1 Introdução

Os Centros de Prevenção e Controle de Doenças estimam que 5% a 15% de pacientes hospitalizados adquirem algum tipo de infecção hospitalar. Apesar dos avanços das técnicas de esterilização e o uso de materiais descartáveis, as taxas de infecções hospitalares aumentaram 36% nos últimos 20 anos. Somente nos Estados Unidos, cerca de dois milhões de pessoas contraem infecções hospitalares por ano e mais de 100 mil morrem como consequência, com um custo anual estimado em 4,5 bilhões de dólares (ANDREI; ZERVOS, 2006).

No Brasil a situação é mais grave, de acordo com os dados do Sistema de Informação Hospitalar do Ministério da Saúde do Brasil (SIH/SUS, 2011). O SIH/SUS reporta que cerca de 8,8% (3.406.521) do total de internações realizadas entre janeiro de 2008 e junho de 2011 possuem relação com algum tipo de doença infecciosa ou parasitária, e 23% (799.029) destas infecções foram causadas por septicemia e outras infecções bacterianas, levando a óbito cerca de 17% (138.227) do total de pacientes contaminados.

A transmissão de microrganismo em ambiente hospitalar pode ser direta, por contato manual ou indireto, por inalação, ingestão ou por penetração na pele. Infecções causadas por microrganismos, com a entrada destes na corrente sanguínea, podem gerar tipos de bacteremia. Segundo Bone et al. (2009) a bacteremia pode ocorrer de forma transitória, intermitente ou persistente, e ser ou não associada à sepse, dependendo do estado imunológico do paciente e do processo infeccioso desencadeante. Muitos fatores podem influenciar à entrada e permanência de microrganismos em pacientes hospitalizados, os quais podem estar mais ou menos susceptíveis a estes agentes (LOUREIRO et al., 2002).

O delineamento correto de linhagens de microrganismos patogênicos é essencial para o estabelecimento das características epidemiológicas destas infecções e no desenho de estratégias de controle eficientes. A identificação genética de mi-

croorganismo em laboratórios e clínicas, é realizada através de métodos com pouca sensibilidade, reprodutibilidade e padronização, além de serem baseados na análise de características fenotípicas. O desenvolvimento da técnica de sequenciamento proporcionou um grande impacto no diagnóstico e controle de muitas doenças infecciosas (DEGENER et al., 1994).

Esses métodos de identificação de microrganismos por sequenciamento são baseados em técnicas utilizadas para o estudo das relações evolutivas, populacionais e da ecologia microbiana através de regiões gênicas específicas ou com o próprio genoma completo (CLARRIDGE, 2004).

O produto do sequenciamento, ou seja, a sequência de DNA, quando aplicado à identificação de microrganismos fornecem resultados mais padronizados (PATEL, 2001). As sequências de ácidos nucleicos obrigatoriamente são depositadas em bancos de dados públicos, e muitos deles possuem controle de qualidade para essas sequências, conhecidos como bancos de dados curados (MIGNARD; FLANDROIS, 2006).

A disponibilidade dessas informações é considerada um fator de importância para o monitoramento e identificação de microrganismos patogênicos, em especial bactérias, pois possibilitam, além de uma identificação mais precisa o acompanhamento de agentes patogênicos emergentes (URWIN; MAIDEN, 2003; COOPER; FEIL, 2004; FEIL et al., 2004).

Considerando a dinâmica que ocorre em ambientes insalubres, como os hospitais, as informações fornecidas pelas sequências de ácidos nucleicos, comparados com outros métodos de identificação molecular de microrganismos proporcionam informações sobre relações e dinâmicas genealógicas de forma direta (CLOUD et al., 2004).

A determinação taxonômica correta dentro do gênero é um fator essencial para que se possa chegar a generalizações e conclusões acerca do relacionamento de qualquer microrganismo, uma vez que a falta de linhagens bem definidas causam grandes equívocos na comparação de diversos casos clínicos estudados (CLARRIDGE, 2004).

Assim, esta pesquisa priorizou as principais técnicas que são utilizadas e validadas em bioinformática, bem como os métodos de análise e armazenamento de dados para identificação de microrganismos por meio do gene rRNA 16S, no

contexto do sistema de saúde pública da Unidade de Emergência Dr. Daniel Houly, localizada na cidade de Arapiraca, situada no interior do Estado de Alagoas.

Espera-se com o desenvolvimento desse sistema, a adequação de ferramentas e métodos consolidados em bioinformática, aplicados as informações produzidas sobre os microrganismos, em especial bactérias patogênicas, implementadas em sistemas de informação e gestão aplicado à saúde pública, com vistas a melhoria do controle da dispersão, clonalidade e registro da resistência a antimicrobianos.

2 Fundamentação Teórica

2.1 Infecção hospitalar

As infecções adquiridas em hospitais são definidas como aquelas que ocorrem em pacientes hospitalizados que não estavam presentes ou incubada no momento da admissão. Esta definição tem como objetivo de diferenciar infecções associadas a internação em hospitais, como consequência da exposição ao ambiente hospitalar, cuidados à saúde, diagnósticos e intervenções terapêuticas, daquela chamada comunitária. Considerou-se também que a prevenção da hospitalização e a higiene apropriada seria capaz de prevenir a sua ocorrência (LACERDA; JOUCLAS; EGRY, 1996).

A definição de Infecção Hospitalar, no Brasil, é regulamentada através da Portaria 2.616/1998 do Ministério da Saúde, que diferencia a Infecção Comunitária (IC) da Infecção Hospitalar (IH), como segue:

Infecção comunitária

- É aquela constatada ou em incubação no ato de admissão do paciente, desde que não relacionada com internação anterior no mesmo hospital;
- Quando associada com complicação ou extensão da infecção já presente na admissão, a menos que haja troca de microrganismos com sinais ou sintomas fortemente sugestivos da aquisição de nova infecção;
- Acontece em recém-nascido, cuja aquisição por via transplacentária é conhecida ou foi comprovada e que tornou-se evidente logo após o nascimento (exemplo: herpes simples, toxoplasmose, rubéola, citomegalovírus, sífilis e AIDS);
- Quando associada a recém-nascidos, cuja mãe permaneceu com bolsa rota superior a 24 (vinte e quatro) horas.

Infecção hospitalar

- É aquela adquirida após a admissão do paciente no ambiente hospitalar e que se manifeste, durante a internação ou após a alta, também quando puder ser relacionada com a internação ou a procedimentos hospitalares.

Além dessas definições, a IH também é classificada de acordo com sua manifestação (doenças infecciosas e síndrome infecciosa) e a topografia (urinárias, acesso vascular, sítio cirúrgico, pneumonias), sendo essa última a mais utilizada pelas Comissões de Controle de Infecção Hospitalar (CCIH) (LACERDA; JOUCLAS; EGRY, 1996).

Com o passar dos anos, se tornou evidente que a taxa de infecção pode ser reduzida com medidas de controle, maior cuidado e acompanhamento, como se observa em Unidades de Terapia Intensiva (UTI), local no qual o paciente se encontra vulnerável pela sua condição de saúde e sendo o setor com um maior número de procedimentos invasivos, no qual observam-se a maior densidade de IH (LACERDA; JOUCLAS; EGRY, 1996; GIUNTA; LACERDA, 2006).

Há vários fatores que levam a patologias devido a infecção, e são parcialmente passíveis de medidas de controle, considerados os principais determinantes de IH. O mais notável entre os fatores encontra-se na gravidade da doença do paciente, tipo e duração dos procedimentos cirúrgicos e a taxa de utilização de dispositivos e procedimentos invasivos. E no lado do patógeno, a seletividade de características de saúde diversas, tais como virulência, propriedades de adesão e a resistência aos antibióticos (GIUNTA; LACERDA, 2006).

As bactérias responsáveis por causar IH são originárias da flora endógena dos próprios pacientes, endógena inespecífica e multirresistente (flora do paciente com resistência adquirida) ou de fontes exógenas. Em contraste com as infecções endógenas, infecções exógenas são mais susceptíveis de serem prevenidas, com medidas de precauções específicas concebidas para reduzir as taxas globais de transmissão nosocomial (LACERDA; JOUCLAS; EGRY, 1996).

Assim, dentro do processo da manifestação de IH, existe a necessidade de entender a cadeia epidemiológica da IH, no qual o paciente encontra-se envolvido no processo infeccioso e nos reservatórios, para um possível hospedeiro vulnerável.

O objetivo principal do controle de infecção é quebrar a cadeia de infecção

(Figura 2). Essa cadeia apresenta seis ligações, e todas precisam estar intactas para que o processo de disseminação de uma doença infecciosa ocorra (SEHULSTER et al., 2003).

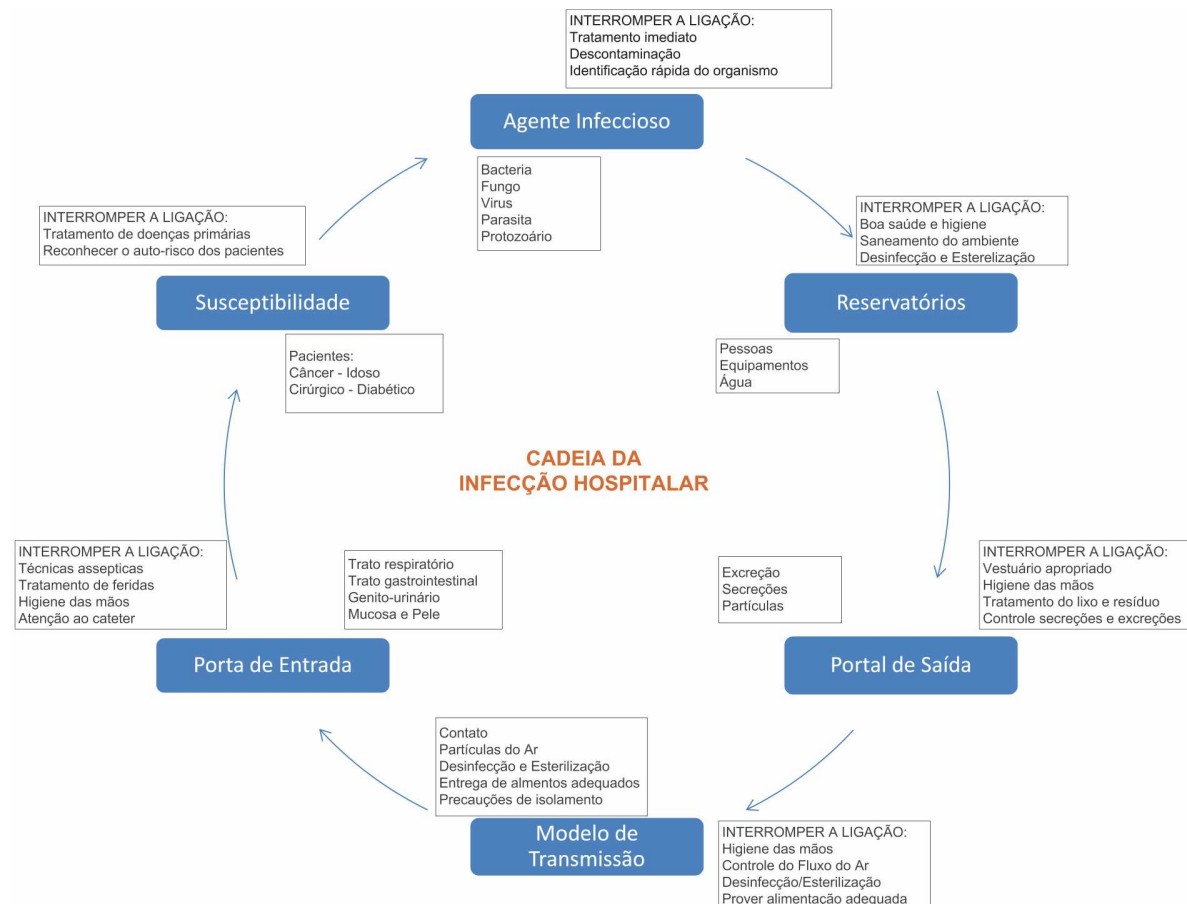


Figura 2.1: Cadeia de transmissão de infecção hospitalar. Modificado de SEHULSTER et al. (2003).

Assim, para quebrar as ligações da cadeia de transmissão de patógenos no ambiente hospitalar, devem ser observados os seguintes elos de ligação (SEHULSTER et al., 2003):

Agente infeccioso: é um microrganismo. Maneiras de eliminar o microrganismos incluem a rápida identificação, com o pronto tratamento da infecção. Outros métodos que eliminam os microrganismos incluem a prevenção de contaminação cruzada e a remoção física de substâncias contaminadas.

Reservatório: uma "moradia", no qual o microrganismo pode sobreviver. Os reservatórios mais comuns associados com infecções adquiridas no hospital são pacientes, profissionais de saúde, equipamentos de saúde e meio ambiente.

Portal de saída: um caminho para os microrganismos saírem de sua "moradia".

Portais de saída comumente associados aos reservatórios humanos incluem tanto o respiratório, genito-urinário, além da pele e da membrana mucosa, sangue, e através da placenta.

Modo de transmissão: o método de transporte que os microrganismos usam para começar a dispersar de sua "moradia" para uma nova "moradia". Essa transmissão pode ser realizada através de quatro meios, que são: por (a) com contato direto com a fonte, contato indireto, e através de um objetivo intermediário, (b) veículo comum (quando os agentes infecciosos estão em um "meio", tal como sangue, comida ou água), (c) no ar, e (d) por vetores (insetos e suas mordidas).

Porta de entrada: caminho aberto para os microrganismos entrarem em uma nova "moradia". Portas de entrada para o trato respiratório, genito-urinário, e tratos gastrointestinais; através da pele ou das membranas mucosas; transplacentária (de feto para a mãe), e parenteral (por sangue através de punções de agulha).

Hospedeiro susceptível: uma nova "moradia" para os microrganismos. Um hospedeiro adequado é um ser humano que é incapaz de resistir a infecção pelo agente infeccioso. Fatores que afetam a suscetibilidade do hospedeiro relacionada com a gravidade de uma infecção incluem a idade, o sexo, a etnia, e os antecedentes de doenças ou doenças subjacentes, o estado nutricional, o comprometimento do estado imunológico, e o estado de trauma.

Assim, a IH ocorre em decorrência de possibilidades de invasão, multiplicação e ação do agente infeccioso no hospedeiro, ocorrendo interação imunológica, com presença de sinais e sintomas que caracterizam a doença ou uma síndrome infecciosa.

2.2 Biologia de bactérias

2.2.1 Genômica de Bactéria

O avanço dos métodos e técnicas em biologia molecular, genética e da bioinformática, permitiram que o primeiro genoma de bactéria fosse sequenciado em 1995, o do *Haemophilus influenzae* (FLEISCHMANN, 1995). Atualmente existem 1715 genomas de bactéria completamente sequenciados, mais de 90 projetos geno-

mas de Archea e 5140 de Bacteria em alguma fase de finalização (PAGANI et al., 2012).

O crescimento dos estudos genômicos dos microrganismos foram impulsionados para se ter uma melhor compreensão da biologia e da evolução das suas relações, bem como para a identificação de genes de virulência, e alvos para os novos mecanismos terapêuticos, como os antimicrobianos (PAGANI et al., 2012).

A possibilidade de análise desses dados somente foi possível devido a evolução dos programas e algoritmos desenvolvidos para a recuperação dessas informações. Esses programas vem, ao longo das duas últimas décadas, melhorando a capacidade de análise, consultas e a recuperação dessa informação em grandes bases de dados biológicos (GLASS et al., 2006; BRAZAS et al., 2011; GALPERIN; FERNANDEZ-SUAREZ, 2012).

Para se ter acesso a essas informações genômicas a partir do sequenciamento do genoma dos microrganismos, as sequências de DNA são então comparadas com sequências de banco de dados públicos (curados ou não curados) usando sistemas de recuperação de sequências. Assim, para identificação ou outro estudo de um microrganismo patogênico, é necessário ter uma boa compreensão da relação entre os genes, somente possível através do acesso à bancos de dados biológicos (ZDOBNOV et al., 2002; HARTE et al., 2004).

2.2.2 Taxonomia e evolução de bactérias

A taxonomia de microrganismos encontra-se em processo de mudança, principalmente devido ao progresso tecnológico, ocasionando alterações metodológicas na identificação de novos organismos. Com o passar dos anos ocorreu um grande aumento na disponibilidade de novos dados, e esse aumento somente foi possível devido ao desenvolvimento da bioinformática, pois atualmente os dados estão disponibilizados em diferentes bases de dados (BRUNAK et al., 2002).

A identificação e diferenciação de microrganismos, principalmente quando possuem características compartilhadas, são atividades difíceis quando se usam apenas testes bioquímicos (PATEL et al., 1998). Identificando-se por estes métodos não se pode verificar a verdadeira relação taxonômica, ou a realização de estudos de ordem evolutiva de forma elucidativa (PETTI, 2007).

As pesquisas de Woese et al. (1975) e colaboradores, proporcionaram o surgimento de diversas linhas de pesquisas ao redor do mundo, com o uso do gene rRNA 16S como principal ferramenta de identificação molecular. Estudos sobre biodiversidade, metagenômica e de identificação de espécies de difícil cultivo em meios controlados, como os clínicos (TURENNE et al., 2001; BANSAL, 2005; CHRISTEN, 2008), levaram ao aparecimento e desenvolvimento de bancos de dados biológicos públicos curados para o gene rRNA 16S (OLSEN et al., 1992; OLSEN; WOESE, 1993).

2.2.3 Identificação molecular de microrganismos através do rRNA e sua importância

Uma das principais estratégias de diagnóstico de agentes responsáveis por infecções hospitalares consiste na amplificação através da Reação em Cadeia de Polimerase (PCR), seguido do sequenciamento e análise da região conservada codificadora de RNA ribossomal 16S, o qual possibilita uma identificação rápida e independente de conhecimento prévio da espécie de microrganismo em estudo. Esse método comparado com outras técnicas moleculares, requer menores riscos no processo de análise e interpretação dos resultados, reduzindo as possibilidades de falso negativo ou positivo (DEGENER et al., 1994; GATSELIS et al., 2006; ANDREI; ZERVOS, 2006).

A análise da sequência do gene do rRNA 16S conduz ao estabelecimento de padrões de transmissão epidemiológica de bactérias patogênicas no ambiente hospitalar. Esse método é utilizado ainda em identificações de difícil manipulação em meios de cultura, permitindo a análise direta de amostras biológicas e a independência com relação aos meios de crescimento (CLARRIDGE, 2004).

O rRNA 16S também é utilizado com finalidades de identificação filogenética de bactérias. As análises das sequências do rRNA 16S têm se tornado um método padrão para estudos filogenéticos de microrganismos (PATEL et al., 2000).

O genes do RNA ribossômico são caracterizados por serem bem mais conservados, ocorrerem em todas as bactérias, e ainda possuir variações na sequência de uma espécie para outra (MARTÍNEZ-MURCIA; ANTÓN; RODRÍGUEZ-VALERA, 1999; CHAKRAVORTY et al., 2007; CHRISTEN, 2008). Essas regiões variáveis são bem mais comuns no rRNA 16S, e não só possuem o poder de distinguir entre

espécies, como também fornecem resultados em graus de diferença.

A primeira descrição do rRNA 16S foi realizada para a bactéria *Escherichia coli*. Para esta bactéria o gene rRNA 16S possui uma massa de $2,5 \times 10^6 Da$ e um coeficiente de sedimentação de Svedberg de 70S. O ribossomo 70S é constituído de duas subunidades diferentes em tamanho, com valores de sedimentação de 30S (subunidade menor, *small subunit* (LSU)) e 50S (subunidade maior, *large subunit* (LSU)). A subunidade pequena (30S) consiste em uma molécula de rRNA 16S e 21 proteínas diferentes, enquanto que a subunidade maior (50S) contém uma de rRNA 5S e outra de rRNA 23S, com 31 proteínas diferentes. O rRNA 16S da *E. coli*, consiste de 1542 nucleotídeos, dos quais 46% estão pareado entre si (YEANG et al., 2007).

As regiões variáveis do gene rRNA 16S não são distribuídas regularmente no gene, além disso o gene é caracterizado por regiões que também são altamente conservadas (BOYER; FLECHTNER; JOHANSEN, 2001) (Figura 2.2). Para se realizar o estudo dessas regiões são necessários inicializadores específicos que amplificam antes das regiões variáveis (GATSELIS et al., 2006). O produto dessa amplificação pode então ser sequenciado. O grau de conservação dessas regiões é tão elevado, que um único par de inicializadores pode ser utilizado na obtenção de produtos de microrganismos diferentes, ou seja, sem saber nenhuma informação sobre o microrganismo em estudo (JENSEN L. B.; WEBSTER; STRAUS, 1993; GATSELIS et al., 2006).

O gene rRNA 16S apresenta nove regiões hipervariáveis (V1-V9), como observamos na Figura 2.2, identificadas em diferentes tipos de microrganismos. As sequências dentro dessas regiões hipervariáveis são os alvos de estudos para o diagnóstico de bactérias patogênicas e estudos filogenéticos. Para se realizar um estudo de identificação, utiliza-se um único par de inicializadores que flanqueiam mais de uma região hipervariável. Segundo Chakravorty et al. (2007), a região V1 é a melhor para estudos de diferenciação entre *Staphylococcus aureus* e *Staphylococcus* sp. coagulase negativo. As regiões V2 e V3 são as mais recomendadas para distinção de todas as espécies bacterianas para o nível de gênero, exceto alguns gêneros de *Enterobacteriaceae* que são intimamente relacionadas, como as espécies de *Mycobacterium fortuitum*, *Nocardia* sp., *Propionibacterium acnes* e *Rhodococcus equi*. A região V2 distingue entre as espécies de micobactérias e a V3 entre as espécies de

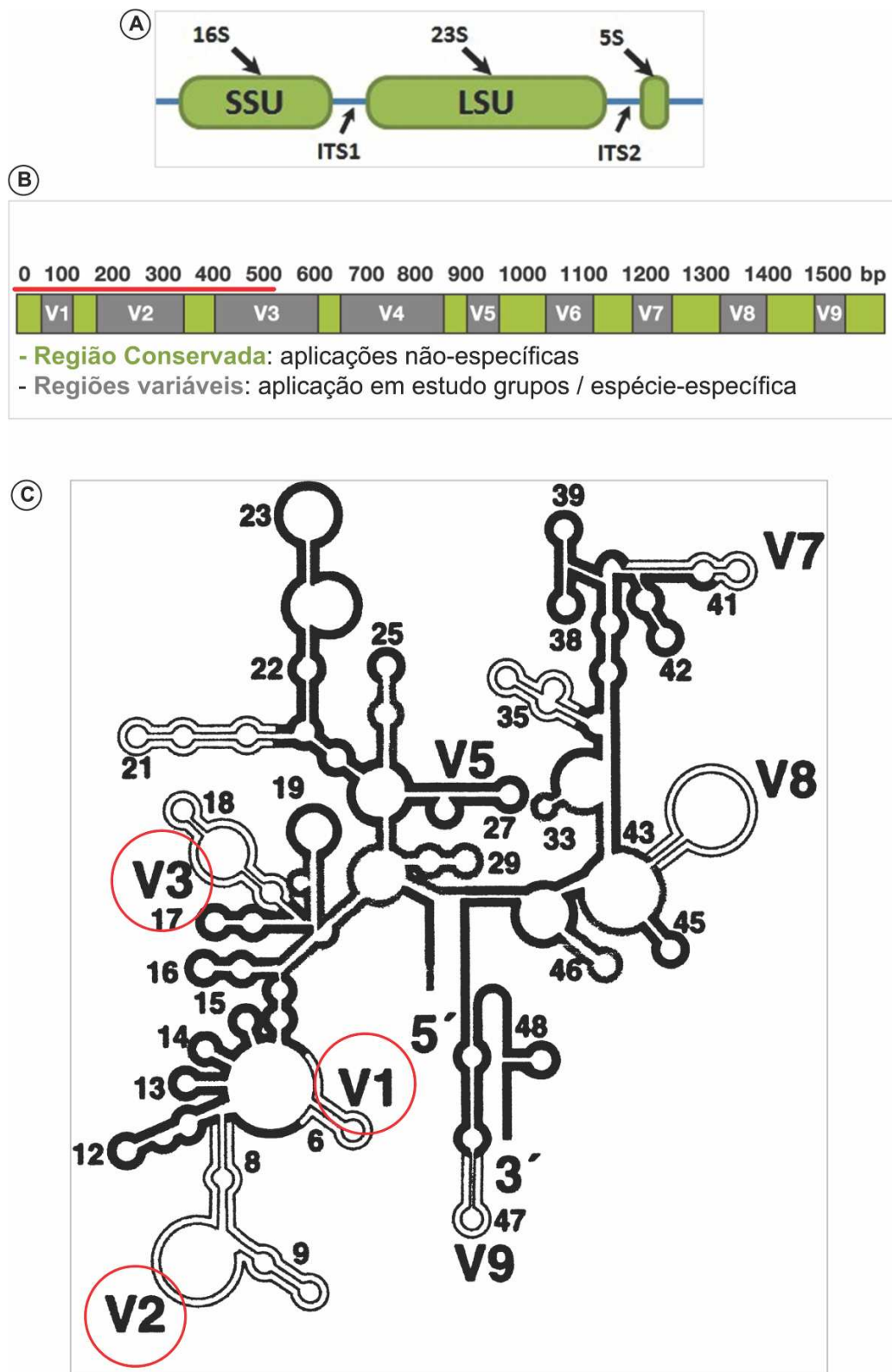


Figura 2.2: O gene rRNA 16S. A - relação entre o rRNA 30S (subunidade menor, *small subunit* (LSU)) e 50S (subunidade maior, *large subunit* (LSU)), bem como as regiões inter-gênicas ITS. B e C - As nove regiões variáveis e o destaque para as regiões V1, V2 e V3 (modificado Yeang et al. (2007))

Haemophilus. Ainda existe um pequeno trecho de sequências de rRNA de aproximadamente 58 pares de bases da região V6, que são eficientes na diferenciação de muitas bactérias, com exceção das enterobactérias. Quanto as regiões V4, V5, V7, V8 e V9 Chakravorty et al. (2007) concluiu que são alvos pouco úteis para estudos de gêneros e espécies.

Muitas discussões tem ocorrido sistematicamente, sobre a justificativa de se utilizar um único marcador molecular para identificação e inferência filogenética de microrganismos fazendo atribuição taxonômica com base nesta filogenia (PACE, 1997; IMHOFF; CAUMETTE, 2004; DAUGA; DORÉ; SGHIR, 2005).

A análise comparativa de sequências rRNA 16S evoluiu, de uma técnica realizada por especialistas e com altos custos, para procedimentos acessíveis em clínicas e laboratórios de microbiologia, favorecendo a identificação de microrganismos, proporcionando assim, uma transformação na taxonomia microbiana atual (WOO et al., 2009).

Os resultados das descoberta de novos genes como marcadores moleculares alternativos, indicam diferenças nos resultados obtidos para a análise filogenética comparativa (CLOUD et al., 2004; FIELDS et al., 2006; CASE et al., 2007; RAS-TOGI et al., 2009). Com o resultado desses projetos, conclui-se que atualmente existem apenas um pequeno conjunto de genes que satisfaçam as condições de marcadores filogenéticos universais, e que estão no núcleo conservado do genoma. Entre eles podemos citar genes além do rRNA 16S como o *recA*, *gyB*, *pyrH*, *rctB*, *rpoD* e *toxR* (ALLERS; MEVARECH, 2005; SCHLEIFER, 2009).

Um dos grandes desafios e divergências quanto a esses estudos, encontram-se em bases de dados que possuem baixa densidade populacional dos genes alternativos, dificultando uma análise mais representativa (CLARRIDGE, 2004; CHAKRAVORTY et al., 2007). Como consequência, as abordagens baseadas em rRNA 16S continuam sendo o padrão-ouro para taxonomia, filogenia e identificação de microrganismos (WOO et al., 2008).

A utilização do rRNA 16S somente foi possível, devido ao rápido desenvolvimento do sequenciamento em alto desempenho, o qual contribuiu para o aumento do conjunto de dados biológicos. Com isso, o rRNA ficou sendo considerado o marcador filogenético melhor documentado em relação ao conteúdo de informação e abrangência (WANG et al., 2007; LIU et al., 2008).

A abrangência dos dados disponíveis em diversas bases de dados da taxonomia atual, bem como a história e a análise da filogenia procariótica, levaram ao gene rRNA 16S como o método mais utilizado para comparação e identificação de microrganismos, em detrimento com outros dados obtidos a partir de marcadores em potencial (Kämpfer; Glaeser, 2012).

As discussões sobre novas espécies são baseadas em uma análise comparativa de alinhamento de sequências em banco de dados específicos, como os disponibilizados através do consórcio de estudo de rRNA SILVA (Pruesse et al., 2007). Outras propostas de reconstrução e identificação são também disponibilizados pelo RDP-II (Maidak et al., 1999; Cole et al., 2007).

Atualmente existem uma grande quantidade de sequências de rRNA 16S em bancos de dados específicos, disponíveis publicamente e curados, como veremos na seção 2.3.3 (Turenne et al., 2001; Harmsen et al., 2003; Wuyts; Perriere; Peer, 2004; Griffen et al., 2011).

2.3 Bioinformática na identificação de bactérias

Os métodos para a identificação de microrganismos são baseados em técnicas que foram sendo aperfeiçoadas ao longo dos anos. Essas técnicas utilizam métodos de inferência das relações evolutivas e incluem estudos populacionais através do sequenciamento de regiões específicas, parciais ou até mesmo com o uso de genomas completos (Kämpfer; Glaeser, 2012).

Com o sequenciamento dessas regiões foram realizadas padronizações que possibilitaram o uso como rotina em muitos estudos de ecologia microbiana (Nemergut et al., 2011) e em laboratórios clínicos (Patel, 2001; Lu et al., 2008). As sequências do gene rRNA 16S utilizadas nesses estudos são depositadas em bancos de dados públicos, e muitos desses bancos possuem curadoria e controle de qualidade (Turenne et al., 2001).

A difusão do rRNA 16S do uso como ferramenta, somente foi possível devido a organização de sequências de ácidos nucleicos nesses bancos de dados, pois o fácil acesso permitiu o monitoramento de diversos microrganismos e a identificação de novas espécies (Olsen et al., 1992; Pruesse et al., 2007; Loy et al., 2007).

O detalhamento de muitas dessas informações são utilizadas atualmente em

estudo da dinâmica das IH (ANDREI; ZERVOS, 2006), portanto poderiam ser utilizadas e aplicadas de forma padronizada em ambientes e sistemas de saúde pública, levando para um melhor controle da dispersão, análise da clonalidade, estudos e padronização da resistências dessas bactérias a antimicrobianos (TANG; PROCOP; PERSING, 1997; STRUELENS; DENIS; RODRIGUEZ-VILLALOBOS, 2004; SINGH et al., July 2006).

Esse mapeamento somente é possível com o uso de ferramentas de bioinformática. Para isso, encontram-se sempre em desenvolvimento diversas ferramentas e técnicas, como algumas estratégias de recuperação de sequências em bancos de dados, alinhamento múltiplo de sequências e construção e armazenamento de bancos de dados específicos (EDGAR, 2004b; KATO et al., 2005; KANZ et al., 2005; LARKIN et al., 2007; GALPERIN; FERNANDEZ-SUAREZ, 2012).

2.3.1 Alinhamento de sequências

O procedimento de alinhamento de sequências é um método de comparar duas sequências (alinhamento par-a-par) ou mais (alinhamento múltiplo de sequências), com o objetivo de localizar séries individuais ou não, ou até mesmo padrões de informações genômicas que se apresentem na mesma ordem (ALTSCHUL et al., 1997; THOMPSON; PLEWNIAK; POCH, 1999b). Os primeiros estudos que abordaram o alinhamento local foram propostos por Sankoff (1972), embora o procedimento de alinhamento local mais utilizado e estudado e o qual serviu como base para a evolução dessa metodologia foram os estudos propostos por Smith e Waterman (1981), que foi adaptado a partir do algoritmo padrão de Needleman-Wunsch (MURATA, 1990).

A visualização do alinhamento local é realizado com a disposição das duas sequências em uma única linha. As bases idênticas ou não, são colocadas em uma coluna, e a visualização do alinhamento é padronizada de acordo com as similaridades entre os ácidos nucleicos, como mostra a Figura 2.3.

Em um alinhamento ótimo as bases não idênticas e as lacunas são colocadas para trazer tantos caracteres idênticos ou semelhantes quanto possível no registro vertical. Sequências que podem ser prontamente alinhadas desta maneira são classificadas como semelhantes.

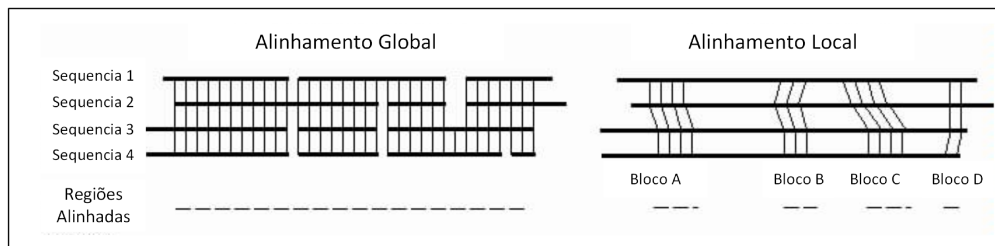


Figura 2.3: Comparação entre o alinhamento global e local (Modificado de Thompson, Plewniak e Poch (1999b)).

No alinhamento local, os trechos da sequência com maior densidade de caracteres são alinhados, gerando uma ou mais ilhas de sub-alinhamento nas sequências (Figura 2.3). Alinhamentos locais são mais adequados para alinhar as sequências que são semelhantes, mas diferentes em outros locais, sequências que diferem em comprimento, ou sequências que partilham uma ou mais regiões conservadas ou domínios (Figura 2.4).

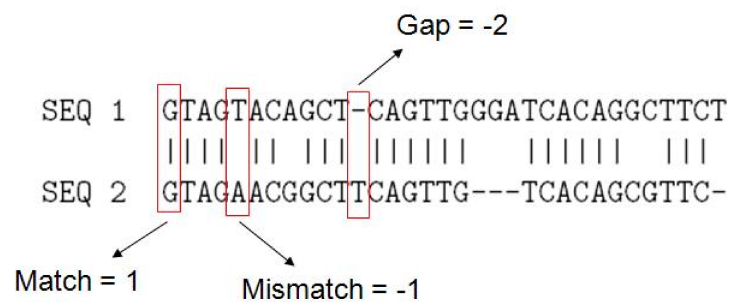


Figura 2.4: Pontuação do alinhamento local (Thompson, Plewniak e Poch (1999b)).

O processo de comparação de sequências que apresentam padrões semelhantes exige de alguma forma a aplicação do alinhamento local. Esse método é a base de todo o sistema de recuperação de sequências em bancos de dados, através de algoritmos, heurísticas e programação dinâmica. O algoritmo de recuperação de sequências mas conhecido é o BLAST (ferramenta para busca de através do alinhamento local) (ALTSCHUL et al., 1990).

Nesse tipo considera-se alinhamentos de subseqções sem a possibilidade de referencias a padrões globais das sequências que estão sendo estudadas. Com essa prerrogativa o algoritmo realiza o alinhamento de regiões de forma separada e independente do padrão global da sequência possibilitando o alinhamento de regiões semelhantes, e ao mesmo tempo permitindo que regiões que possuem características altamente divergentes permaneçam desalinhadas (ALTSCHUL et al., 1990; ALTSCHUL et al., 1997).

No alinhamento global é realizada uma tentativa para alinhar toda a sequência, até ambas as extremidades de cada sequência. Sequências que são bastante semelhantes e aproximadamente do mesmo comprimento são candidatas adequadas para o alinhamento global (Figura 2.5).

O alinhamento global é um tipo de alinhamento que considera que as sequências possuem um alto nível de homologia e o algoritmo alinha todas as sequências da melhor forma possível, identificando os locais com maior identidade. Assim, quando ocorre diferenciações localizadas surgem outras abordagem que podem ser utilizadas na substituição do alinhamento global, o alinhamento local (THOMPSON; PLEWNIAK; POCH, 1999b; BAXEVANIS; OUELLETTE, 2001).

Portanto, o desenvolvimento de algoritmos de alinhamento foram impulsionados devido ao crescimento das sequências que eram armazenadas em bancos de dados biológicos. Grupos de pesquisas avaliaram que o custo computacional era alto e que os estudos com ênfase em algoritmos se faziam necessários para recuperar sequências de forma mais otimizada (ZDOBNOV et al., 2002; KANZ et al., 2005).

2.3.2 Alinhamento múltiplo de sequências

O avanço na obtenção de novas sequências de DNA levaram também ao desenvolvimento de novos métodos de alinhamento múltiplo de sequências. A análise desses dados tem aumentado, devido às buscas das relações entre os diversos organismos existentes (HERTZ; STORMO, 1999; NOTREDAME, 2002).

O processo e a análise dos dados de um alinhamento múltiplo de sequências é um procedimento de rotina em bioinformática. Muitos métodos utilizados em diversos estudos ainda não avançaram no entendimento das informações biológicas e superação do custo computacional para garantir que um alinhamento biológico seja perfeito (SPANG; VINGRON, 2001).

A metodologia padrão utilizada para a identificação de bactérias patogênicas, consiste no estudo da amplificação da região que codifica para o gene rRNA 16S e seu posterior sequenciamento, e em seguida, análise através de alinhamento múltiplo de sequências em busca de regiões homólogas. Assim, o alinhamento de sequências é a primeira etapa da análise em bioinformática (WOO et al., 2009).

Considera-se que as sequências de interesse possuam relações de ancestrali-

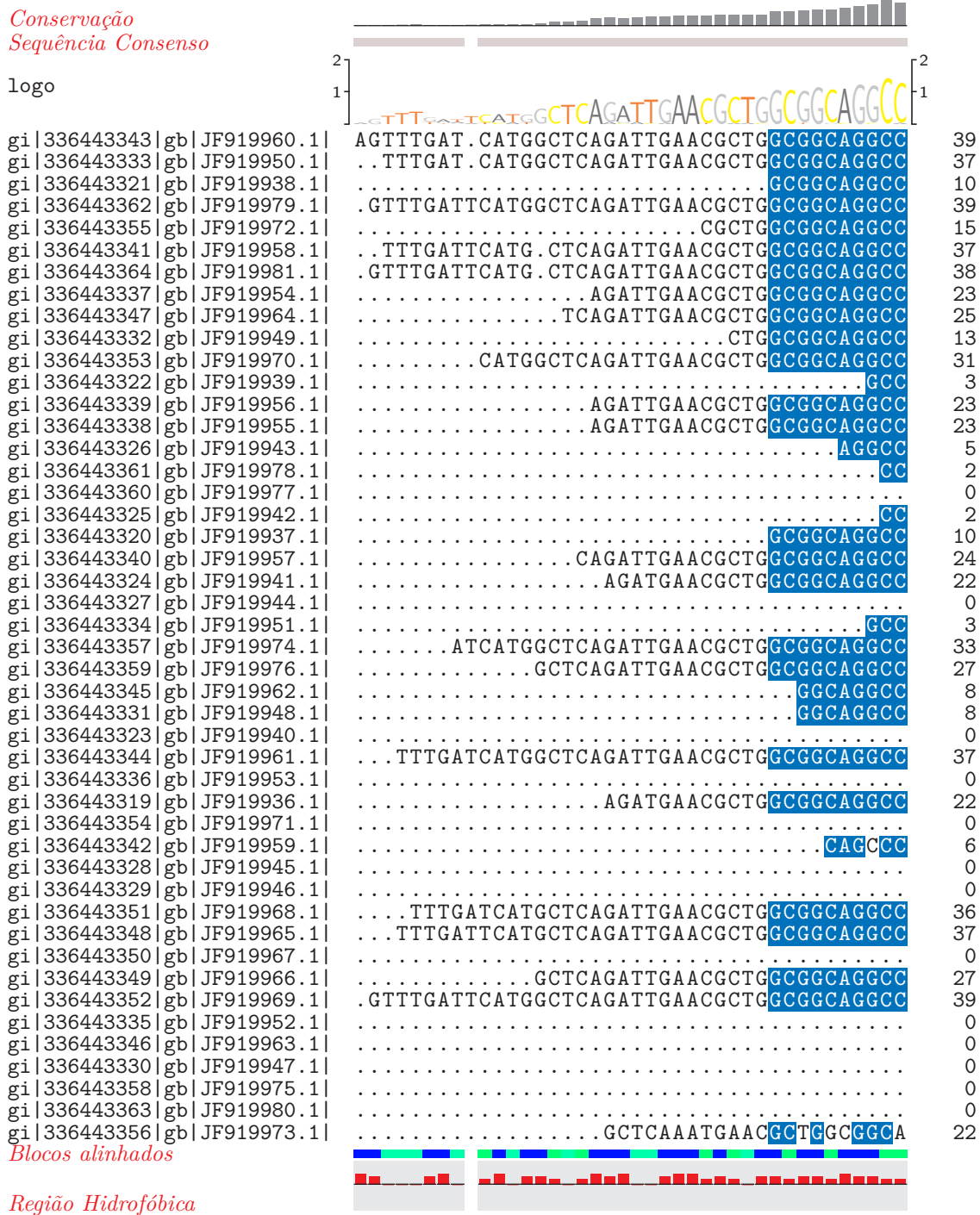


Figura 2.5: Alinhamento múltiplo de sequências de *P. aeruginosa*. Autor, 2012.

dade, assim, busca-se pontos de homologia de sequências que são considerados e avaliadas durante o processo de alinhamento. O resultado do alinhamento múltiplo de sequências considera hipóteses de homologia para cada nucleotídeo do gene rRNA 16S (PATEL, 2001; CLARRIDGE, 2004; WOO et al., 2009).

No estudo de identificação de microrganismos patogênicos busca-se relações entre espécies distintas e relacionadas, que são estimadas através de regiões conservadas, como os próprios genes ribossomais (MAHENTHIRALINGAM; BALDWIN; DOWSON, 2008).

Assim, para uma análise computacional mais informativa se faz necessário que os estudos e fluxos possam proporcionar um intercâmbio de comunicação com os programas de alinhamento de sequências, editores e visualizadores de alinhamento específicos (DEVULDER et al., 2003).

Atualmente existem diversos programas de edição que são capazes de realizar esse tipo de conversão, fazendo com que o resultado possa ser analisado por outros programas de estudos filogenéticos. Entre eles podemos citar o Jalview (WATERHOUSE et al., 2009), MEGA (TAMURA et al., 2011) e Archaeopteryx (TAMURA et al., 2011).

Descrevemos a seguir os principais métodos de alinhamento múltiplo que são utilizados, como os métodos progressivos (FENG; DOOLITTLE, 1987), aqueles baseados em consistência (MORGENSTERN; DRESS; WERNER, 1996) e os iterativos (HIROSAWA et al., 1995; THOMPSON; PLEWNIK; POCH, 1999b).

Métodos progressivos: Uma característica dos algoritmos de alinhamento progressivo está na eficiência em seus resultados, e também no tempo de resposta mais rápido do que outros algoritmos. Entre os programas mais utilizados, e que são classificados como progressivos, podemos citar o CLUSTALW (THOMPSON; HIGGINS; GIBSON, 1994). Thompson, Plewniak e Poch (1999b) realizaram teste para avaliação do desempenho do alinhamento múltiplo de sequências com base nos dados de referência de alinhamentos estruturais, e mostrou que o CLUSTALW apresenta resultados relativamente mais precisos (THOMPSON; PLEWNIK; POCH, 1999a).

Métodos baseados em consistências: Os programas de alinhamento múltiplo são caracterizados em sua grande maioria como algoritmos progressivos (DO et al., 2005; EDGAR, 2004a; KATO; TOH, 2008; NOTREDAME, 2002; PEI; GRISHIN, 2007; THOMPSON; HIGGINS; GIBSON, 1994). O método progressivo apresenta, no início do algoritmo, um pequeno erro de processo. Muitos programas foram adaptados e aperfeiçoados, em relação aos métodos até então utilizados (GOTOH,

1999; WALLACE; BLACKSHIELDS; HIGGINS, 2005). O método progressivo utiliza a pontuação do alinhamento simples entre duas sequências, que provavelmente é a parte central da estratégia progressiva.

Com o avanço nos estudos de algoritmos para alinhamento múltiplo de sequências, foram desenvolvidas estratégias que melhoraram esse tipo de abordagem. A melhoria da estratégia do alinhamento progressivo foi implementado no pacote de software T-COFFE (NOTREDAME, 2002), onde as sequências passam por um processo de alinhamento de forma progressiva. Para isso, utilizam consistência baseada em uma função que reduz possíveis erros nas primeiras fases da construção do alinhamento, o que não ocorre na abordagem do alinhamento progressivo. A melhoria realizada no método consistente, cria duas bibliotecas primárias geradas a partir dos alinhamentos iniciais, verificando posteriormente a consistência do mesmo; em seguida atribui pesos e conclui utilizando a abordagem progressiva.

Métodos iterativos: O aprimoramento dos métodos de alinhamento múltiplo levam sempre em consideração as propostas e resultados obtidos em outros métodos. A estratégia do método iterativo é produzir um alinhamento usando as técnicas do alinhamento progressivo e em um segundo momento incrementar os resultados através de uma série de iterações, ou ciclos, até que os resultados obtidos não possam ser mais melhorados (HIROSAWA et al., 1995; THOMPSON; PLEWNIAK; POCH, 1999b). O método iterativo foi implementado através dos algoritmos de dois programas bem conhecidos, o MUSCLE (EDGAR, 2004b) e o MAFFT (KATO et al., 2005), o último considerado o algoritmo atualmente mais rápido.

Assim, as análises realizadas por Notredame (2002), Edgar (2004a), e Katoh et al. (2005) compararam o desempenho desses programas, usando bancos de dados de alinhamento múltiplo de referência como BALiBASE, e chegaram a alguns resultados divididos em categorias de velocidade e precisão da análise. Observando a primeira característica, comprovou-se a seguinte ordem do mais veloz para o menos: MUSCLE, MAFFT, CLUSTALW e T-COFFEE, com relação a precisão foi o MAFFT, MUSCLE, T-COFFE E CLUSTALW.

2.3.3 Bancos de dados biológicos

A publicação da *Nucleic Acids Research Database* (NAR 2012) registrou 1380 bancos de dados validados, e apresenta 92 novos bancos de dados que disponibilizaram informações genômicas. Além disso, esses novos bancos de dados apresentaram novas ferramentas para análise de informações (GALPERIN; FERNANDEZ-SUAREZ, 2012). Mais de 100 artigos descreveram algum tipo de atualização em bancos de dados já publicados pela NAR Database. Esse aumento reflete na disponibilidade do número de sequências genômicas que foram geradas a partir de novas abordagens e técnicas experimentais e os avanços no estudo dos processos celulares e moleculares.

Com a quantidade de dados gerados devido aos estudos genômicos, os principais bancos de dados cresceram substancialmente, como podemos observar na Figura 2.6. A sistematização, organização e armazenamento das informações biológicas são gerenciadas por grandes bancos de dados, alguns públicos outros com acesso restrito. Esse crescimento levou a especialização dos bancos de dados, que alguns chamam de bancos de dados biológicos, pois possuem dados e informações do campo da genômica, proteômica, metabolômica, farmacogenômica, e de sistemas biológicos (KODAMA et al., 2012; GALPERIN; FERNANDEZ-SUAREZ, 2012).

O uso desses bancos de dados encontra-se em crescente acessão e atualmente vem sendo campo de estudo por pesquisadores das ciências genômicas, refletindo no progresso e avanço da medicina. Mas essas aplicações não evoluíram ao ponto de serem implantadas nos grandes sistemas de saúde, ficando restritas apenas a pesquisas e desenvolvimento com perspectivas de implantação (SIDHU; CHEN, 2008).

Os avanços do diagnóstico e prevenção de doenças, os quais já possuem protocolos envolvendo os dados armazenados em bancos de dados biológicos, como padrões definidos por técnicas moleculares, vem proporcionando ambientes de trabalho mais eficientes (SIDHU; CHEN, 2008).

Esses bancos de dados também proporcionaram avanços significativos na geração de informações filogenéticas, pois o armazenamento de grandes volumes de dados é um dos fatores decisivos para geração de conhecimento. O desenvolvimento desses bancos de dados também levaram ao crescimento do número de ferramentas necessárias para avaliar a quantidade e qualidade da informação armazenada, dire-

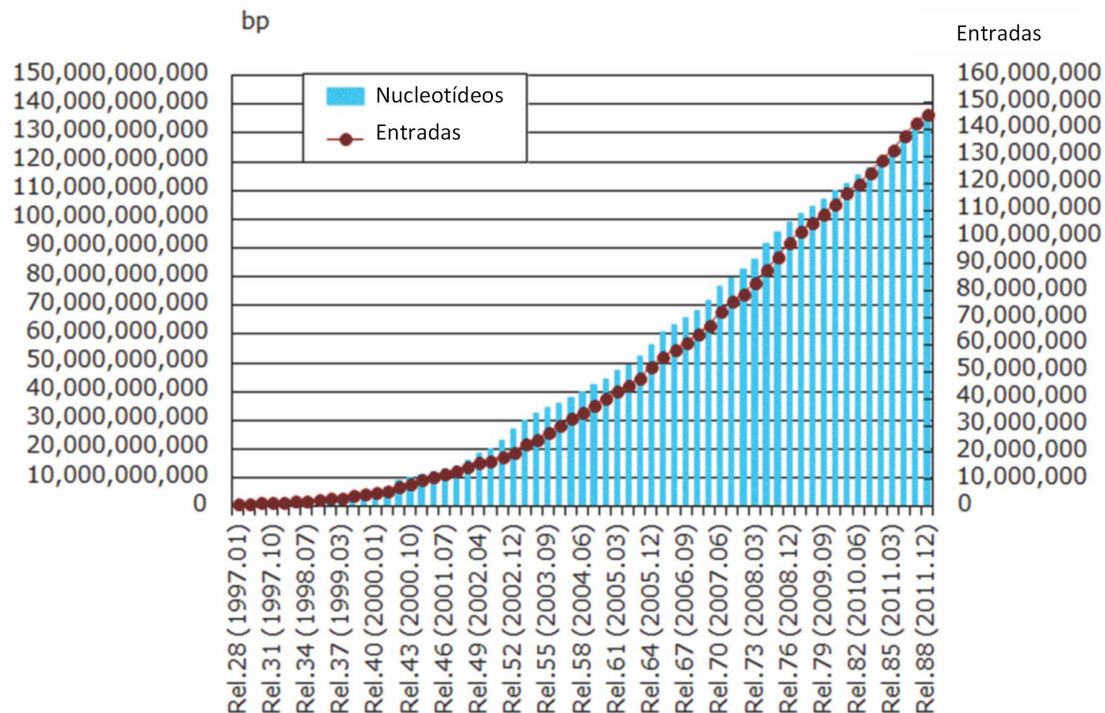


Figura 2.6: Crescimento dos bancos de dados DDBJ/EMBL/GenBank versão 88 (Adaptado de(KODAMA et al., 2012)).

cionando para uma melhor compreensão da reconstrução filogenética e identificando áreas de taxonomia que necessitam de mais estudos, principalmente de um maior número de sequências de regiões genômicas representativas e de interesse para os taxonomistas.

2.3.4 Bancos de dados de rRNA

O sequenciamento do gene que codifica para o rRNA atualmente é considerado como padrão-ouro na identificação de bactérias patogênicas. A popularização desse método teve crescimento devido a disponibilização e acesso livre dos dados, além disso, o depósito de sequências de rRNA em bases de dados especializadas e curadas tem aumentado, como podemos observar na Figura 2.7.

O suporte da bioinformática para a compreensão das doenças infecciosas, vem sendo fundamentado em estudos da genômica microbiana, estudos comparativos e de identificação e caracterização de suas principais propriedades relacionadas com a reconstrução do mapeamento e dispersão dessas infecções (HARMSEN et al., 2002; LIU et al., 2008).

Atualmente existem três grandes projetos de sequenciamento, armazenamento

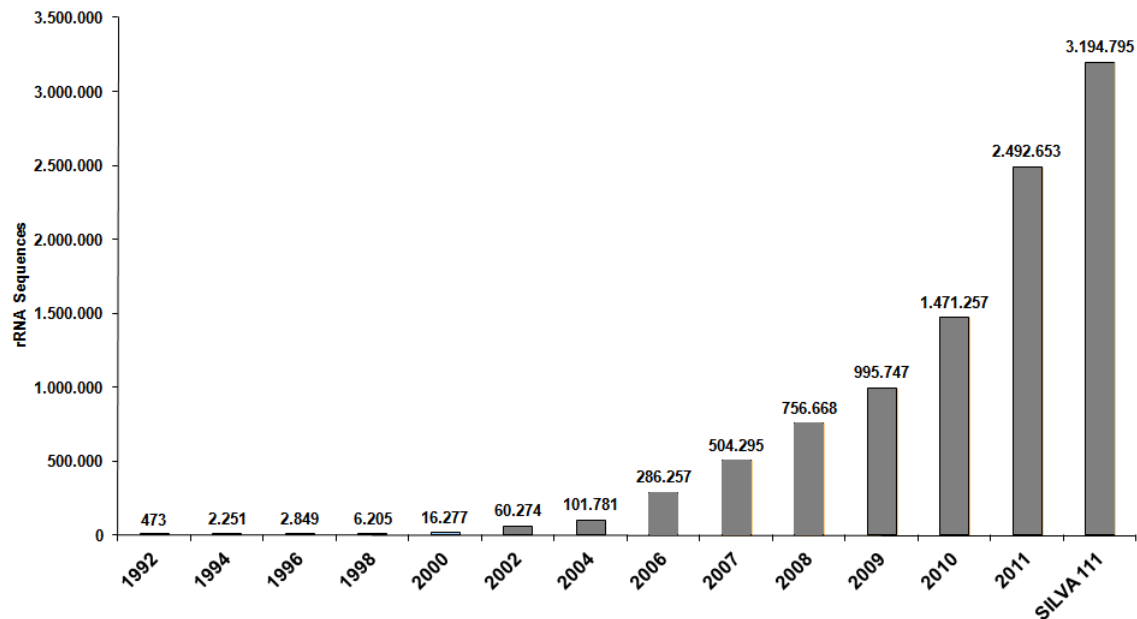


Figura 2.7: Crescimento do banco de dados de rRNAs. (PRUESSE et al., 2007)

e disponibilização de sequências alinhadas ou não, de rRNA 16S através da web, com verificação de qualidade, e em constante atualização, como segue.

Ribosomic Database Project II (RDP-II): mantido pela Michigan State University em East Lansing (<<http://rdp.cme.msu.edu>>), com objetivo principal de armazenar sequências de rRNA de Bacteria e Archaea. O RDP-II possui um sistema de recuperação de sequências. Também possibilita a navegação através de um sistema taxonômico avançado. Disponibiliza ferramentas baseadas na web, como o classificador RDP, busca por similaridade como o *SeqMatch*; prospecção de coincidências através do *Match* e comparação e construção de árvores, que permite uma primeira análise e posicionamento taxonômico de sequências. Possui também ferramentas web interativas para visualização e mapeamento das relações entre muitas sequências ao mesmo tempo (COLE et al., 2005; MAIDAK et al., 2001).

GREENGENES: mantido pelo Laboratório Nacional Lawrence Berkeley em Berkeley (<<http://greengenes.lbl.gov/>>), possui 1.049.116 sequências de rRNA obtidas de Bacteria e Archaea. Uma das particularidade do GREENGENES foi a constituição de um *filo-chip*, com uma grande quantidade de sondas e mais de 45.000 OTUS (Unidades Taxnômicas Operacionais), disponibilizadas através da tecnologia Affymetrix. Combinada com o software *PhyloTrac*, disponibilizado através da web, possibilitando que comunidades microbionas sejam acompanhadas quantitativa-

vamente (DESANTIS et al., 2006).

SILVA: é mantido e organizada pelo Instituto Max Planck de Microbiologia Marinha, em Bremen, Alemanha (<<http://www.arb-silva.de>>). É considerado o maior sistema de análise para rRNA disponibilizado através da web. O banco de dados possui controle de qualidade para as sequências alinhadas de rRNA, de Bacteria, Archaea e Eukarya indexados e compatíveis com o pacote de software ARB. As sequências armazenadas passam por processos de verificação em busca de anomalias em um grande conjunto de dados, e possui sistema de classificação que proporciona ao usuário uma visão da qualidade de sequências. As sequências de referência armazenadas no SILVA são sinalizadas e relacionadas a um organismos cultivado, ou a uma cepa tipo, ou até mesmo a um projeto genoma específico (PRUESSE et al., 2007).

Os três bancos de dados específicos de rRNA descritos acima, possuem uma variedade de sequências do mais alto nível, para estudos e pesquisa de rRNA. Novas tecnologias de *pipelines* e *workflow* de análise estão sendo desenvolvidas e implementadas atualmente. A tendência no estudo de rRNAs são abordagens integrativas, as quais irão proporcionar uma visão abrangente sobre a diversidade microbiana, esclarecendo a função dos microrganismos tanto em ambientes de difícil acesso pelo homem, quanto na dinâmica infecciosa em ambientes hospitalares, e a sua relação com seu meio e a sua distribuição.

2.4 Bioinformática aplicada a saúde

O gene que codifica para o rRNA 16S é utilizado atualmente como relógio molecular, por proporcionar estudos sobre as distâncias evolutivas entre microrganismos ao longo do tempo. Assim, espécies de bactérias que anteriormente tinham sua identificação imprecisa, passaram a ser facilmente identificadas através do sequenciamento parcial ou total do 16S rRNA, tornando possível a determinação do posicionamento filogenético.

A bioinformática aplicada ao estudo da epidemiologia molecular de microrganismo com uso de identificadores moleculares tem potencial para ser utilizado como rotina no diagnóstico de microrganismos. A bioinformática translacional, como é conhecida atualmente, é uma área que permite uma maior eficiência, desempenho e

integração de muitos sistemas diagnósticos. As informações mais precisas e padrões de resultados baseados em dados moleculares mais próximos do contexto biológico real, tem como objetivo o de otimizar e melhorar a segurança dos pacientes e dos ambientes hospitalares (BUTTE, 2008; KANN, 2010; SARKAR et al., 2011).

A bioinformática translacional une o conhecimento biológico com o clínico, abrangendo os estudos moleculares aplicado à população, formando assim, a base para transposição e aplicação do conhecimento adquirido através da bioinformática (Figura 2.8). O conhecimento das técnicas e processos da bioinformática vem sendo aplicado como núcleo da integração desses dados e o diagnóstico avançado em saúde. Atualmente vem sendo considerado desafio e é de interesse da bioinformática translacional o desenvolvimento e aplicação de sistemas para esse fim (SARKAR et al., 2011).

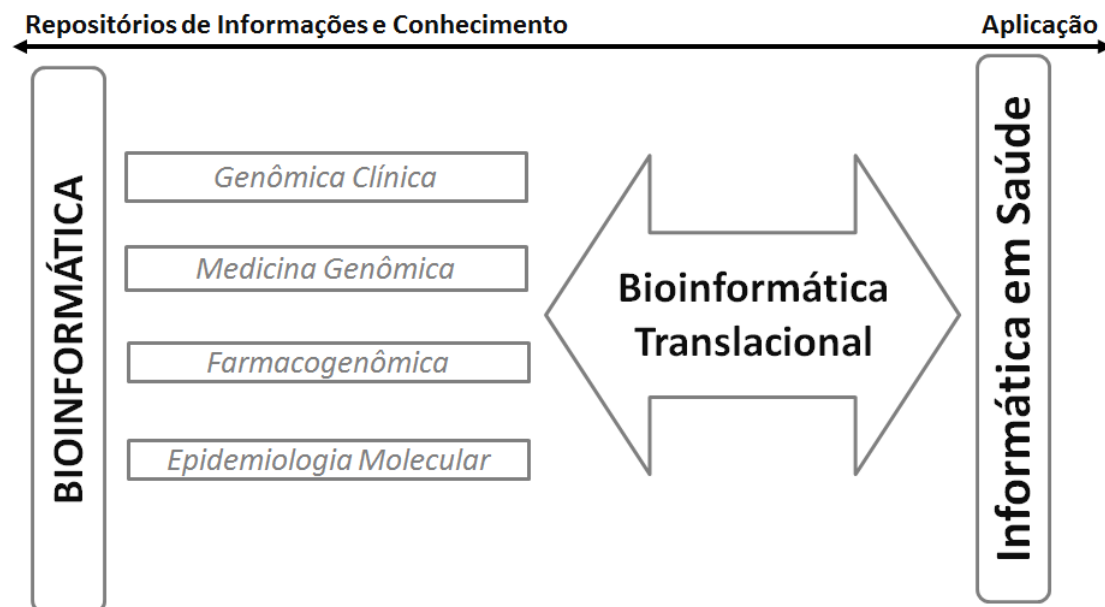


Figura 2.8: Bioinformática translacional união do conhecimento biológico e o clínico. Adaptado de (SARKAR et al., 2011)

2.4.1 Sistema de diagnóstico em saúde

Alguns estudos indicam que erros no diagnóstico microbiológico levam a grande parte da morbidade (SHENG et al., 2005). Nos últimos anos ocorreu aumento em estudos sobre a saúde e o cuidado com paciente, tornando-se uma grande área de estudo (ANGUS et al., 2001). Muitos desses parâmetros servem para analisar e divulgar informações epidemiológicas que avaliam o desempenho dos sistemas

de saúde e promovem meios para alcançar melhores indicadores, ou seja, números mais baixos de morbidade (MCKIBBEN et al., 2006; ROTHBERG; BENJAMIN; LINDENAUER, 2009).

Devemos considerar que as informações sobre a dispersão ou epidemias em hospitais são utilizadas como parâmetros em muitas avaliações de controle de infecções. Atualmente os esforços são concentrados para avaliar o desempenho dos padrões de infecções e são utilizados como medidas de controle de todo o processo (MCKIBBEN et al., 2006).

Os Centros de Prevenção e Controle dos Estados Unidos (CDC) realizaram estudos que estimavam uma taxa anual de infecção de 10% entre os pacientes hospitalizados, e essas infecções estavam associadas ao cuidado inicial com o paciente. As infecções associadas ao cuidado, estão sendo cada vez mais controladas (ANDREI; ZERVOS, 2006; PRONOVOST et al., 2006a; PRONOVOST et al., 2006c; PRONOVOST et al., 2006b).

Atualmente, existem poucos sistemas que são desenvolvidos e modelados com perfis específicos. Esses sistemas são aplicados para o diagnóstico e testes de doenças, podendo ser utilizado para o estudo da comunidade ou para o paciente. Assim, independente do padrão utilizado para o armazenamento das informações e notificações, o desempenho do sistema pode ser avaliado, seguindo os protocolos dos teste de diagnóstico, com perspectivas de sensibilidade e especificidade (GARG et al., 2005; RANDELL et al., 2007; HAYNES; WILCZYNSKI; , 2010).

Os sistemas de apoio a decisão em saúde que disponibilizem mecanismos de notificação e diagnóstico são aqueles considerados os mais eficientes. O seu sucesso está relacionado a formação de uma rede confiável, e responsável pelo fornecimento de dados, através do uso correto de tecnologias seguras utilizadas no gerenciamento e distribuição desses dados para consulta e divulgação (RANDELL et al., 2007; HAYNES; WILCZYNSKI; , 2010).

2.4.2 Sistemas automatizados aplicados à saúde

Atualmente os sistemas de informação aplicados a saúde estão sendo modelados para suportar uma grande quantidade de dados, para atender a demanda da saúde pública, fazendo com que sejam agregados valores as informações obtidas diariamente. O mapeamento de doenças causadas por microrganismos são uns

dos principais objetivos desses sistemas. Um grande desafio atualmente encontra-se no preenchimento do prontuário eletrônico do paciente por completo, com dados da patogenicidade e dos conteúdos necessários para identificação e recuperação da informação posterior para o mesmo paciente. A modelagem desses sistema é o principal caminho para a resolução desses problemas, através de regras confiáveis que devem ser seguidas e aplicadas de forma eficaz e depois reavaliadas de forma a proporcionar resultados mais precisos (KILBREDGE; CLASSEN, 2008; FERNANDO et al., 2012).

O prontuário eletrônico do paciente (PEP) é considerado o foco principal no desenvolvimento de sistemas de informação aplicado a saúde. O PEP refere-se ao armazenamento completo de dados e informações dos registros médicos eletrônico, relacionados com o atendimento, cuidados e exames do paciente. O registros dessas informações proporcionam benefícios para o sistemas de saúde, devido a facilidade das notificações sobre doenças e no produção de relatórios (DEUTSCHER et al., 2008).

Entre as diversas questões que são levadas para sua implantação encontra-se o registro de novas infecções causadas por microrganismos, levando a determinados surtos ou doenças a serem classificadas como endemias, epidemias ou até pandemias emergentes (BROSSETTE et al., 2006).

Em países desenvolvidos tem ocorrido investimentos com a atenção a saúde, e muitos desses investimentos são direcionados ao diagnóstico preciso e na eficácia do tratamento de uma infecção patogênica existente em pacientes hospitalizados de forma individual. A notificação e prevenção dessas infecções tem se apresentado de forma eficaz pois tem ocorrido uma substancial diminuição na carga de infecções comunitárias de uma forma geral, como a tuberculose (KAZUMI; MAEDA; SUGAWARA, 2006) e gripe H1N1 (MONTI; WHITLEY, 2008).

Quando desenvolvido e modelado de forma eficiente os sistemas de informação aplicado a saúde, geram relatórios e estes são as principais fontes de informação para o planejamento de políticas e de ações para a saúde pública. Muitas intervenções realizadas em centros de saúde são norteadas com bases nesses índices, e a tomada decisão para minimizar esses riscos são decisivos no tratamento e intervenção curativa para o sistema de saúde (BROSSETTE et al., 2006)

Os sistemas de vigilância de infecções, tem como principal foco, a busca de

novos casos de doenças infecciosas que podem ter importância para o sistemas de saúde público . As notificações se forem realizadas de forma automática serve para chamar a atenção dos profissionais do serviço e das autoridades em saúde pública. A modelagem de sistemas com essa natureza, com o armazenamento contínuo de dados, podem ser utilizados para intervir de forma mais dinâmica e impedir a propagação ou a contaminação dessas infecções (PITTET, 2005; HOTA et al., 2010).

Assim, a importância do desenvolvimento e modelagem eficaz desses sistemas esta no detalhe dos registro e comunicação da patogenicidade, que são realizados e vinculados a pacientes de forma específica, podendo ser acompanhado ao longo do tempo. Os procedimentos de intervenção eficaz, como o objetivo de se realizar um tratamento profilático ou curativo com antibióticos, são mecanismos adotados como protocolos para minimizar o risco de propagação de doenças (PITTET, 2005; BROSSETTE et al., 2006; HOTA et al., 2010).

3 Objetivos

3.1 Gerais

Desenvolver uma ferramenta de banco de dados para sequências do gene rRNA 16S, visando o armazenamento, análise filogenética e interpretação de dados direcionados para a identificação de bactérias patogênicas, associado a um sistema de busca relacionada aos registro de resistência aos antimicrobianos.

3.2 Específicos

- Modelar e implementar interface de comunicação entre usuário e banco de dados;
- Incorporar no sistema ferramentas de alinhamento, análise de dados moleculares e busca em bases de dados específicas para a região gênica do rRNA 16S;
- Implementar ferramentas para avaliação e comparação da sequência gênica do rRNA 16S depositada no banco de dados local;
- Desenvolver mecanismos de busca e mapeamento de infecção hospitalar causadas por bactérias patogênicas;
- Desenvolver uma interface gráfica para interpretação e análise dos resultados de identificação filogenética, junto a um sistema de recuperação de sequências e sua relação com os resultados dos testes antimicrobianos;
- Desenvolver um sistema de consulta para permitir à análise das sequências obtidas do gene rRNA 16S e estimar a posição filogenética de possíveis cepas de bactérias patogênicas isoladas utilizando o ambiente de banco de dados desenvolvido.

4 Material e Métodos

4.1 Levantamento de requisitos

Para mapeamento dos dados, foi realizado uma estudo de análise de requisitos do sistema, considerada a parte mais importante no ciclo de desenvolvimento de qualquer sistema. O levantamento das necessidades do sistema foi realizado através de entrevistas e mapeamento dos dados contidos nos formulários utilizados para obtenção dos dados (Apêndice A), tanto no prontuário do paciente quando no da Comissão de Controle de Infecção Hospitalar (CCIH), da Unidade de Emergência do Agreste. Além dessas informações, foram obtidos os dados das amostras microbiológicas e moleculares.

Foram também realizados estudos de análise e integração das principais ferramentas em bioinformática e um levantamento dos principais banco de dados para o gene rRNA 16S. O estudo também envolveu o levantamento de requisitos necessários para incorporação de ferramentas padronizadas em bioinformática utilizadas para a recuperação de sequências em bancos de dados, visualização e alinhamento múltiplo de sequências, bem como para estudos filogenéticos.

4.1.1 Desenvolvimento dos sistemas

A interface web do servidor foi desenvolvida utilizando *Cascading Style Sheets* (CSS), que é um linguagem de estilo utilizada para definir a apresentação de documentos, como HTML ou XML. Seu principal benefício é prover a separação entre o formato e o conteúdo de um documento, bem como JavaScript que é uma linguagem utilizada na validação de formulários no lado cliente e proporcionar a interação com a página web no navegador.

Foi utilizado também, o framework JQuery, que separa a parte lógica do sistema da interface web, otimizando os problemas de incompatibilidade dos nave-

gadores atuais. A biblioteca JQueryUI, utiliza os recursos disponíveis na jQuery com o objetivo de abstrair a construção de interações. A JavaServerPages (JSP) baseada na linguagem de programação Java, tem a vantagem da portabilidade de plataforma, que permite a execução do BACSearch e do MolEpi em diversos sistemas operacionais, como o Windows da Microsoft, MacOS da Apple, Unix e Linux além dos principais navegadores.

O Struts é um framework de desenvolvimento da camada controladora, numa estrutura seguindo o padrão Model2 (uma variante do MVC oficializada pela SunMicrosystem) de aplicações web construída em Java para ser utilizado em um *container* web e em um servidor J2EE.

MVC (Model-View-Controller) é um modelo padrão de arquitetura usada em engenharia de software para separar em 3 camadas a regra de negócio, a persistência dos dados e a interface de apresentação ao usuário (Figura 4.1). Sendo um padrão de desenvolvimento, é implementado em várias linguagens de programação e altamente difundida permitindo que equipes de desenvolvedores trabalhem sem interferência, deixando o código mais legível.

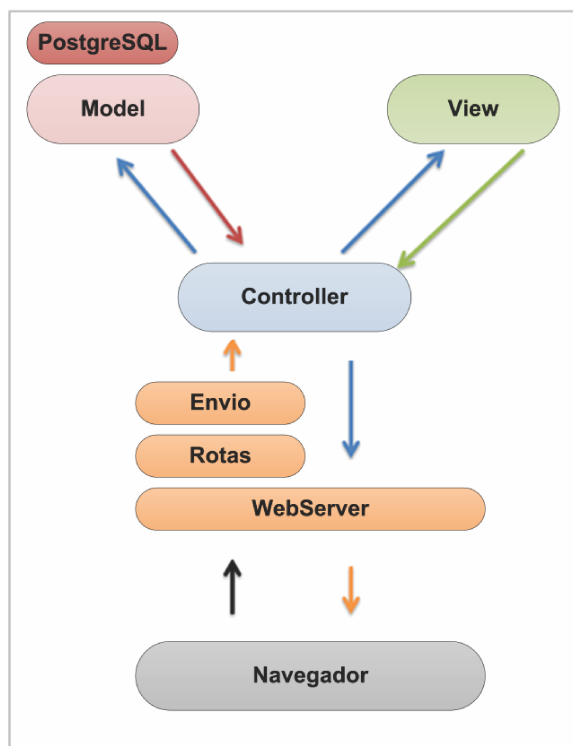


Figura 4.1: O diagrama exemplificando a relação entre Model, View e Controller - MVC, demonstrando as associações diretas e indiretas.

Para o processamento das sequências de consulta utilizamos o framework Bio-

Java. No alinhamento e busca por similaridade utilizamos o QBLAST e o BLASTALL. Todas as aplicações rodam no servidor Apache Tomcat 7.0 (ou superior).

Foi implementado também as ferramentas de recuperação de sequências em banco de dados biológicos, através BLAST (ALTSCHUL et al., 1990), e de alinhamento múltiplo de sequências com o uso dos programas MUSCLE (EDGAR, 2004a), CLUSTALW (THOMPSON; HIGGINS; GIBSON, 1994) e MAFFT (KATO et al., 2005). Os estudos filogenéticos são possíveis através dos visualizadores, analisadores e editores de árvores filogenéticas JalView e Arceuthobium.

4.2 Metodologia de validação

4.2.1 BACSearch

Para validação do BACSearch, foi realizado um levantamento na literatura, de novas espécies bacterianas descritas em 2011. A triagem inicial foi realizada com as palavras-chave (phylogenetic[All Fields] AND 16S[All Fields]) AND "sp. nov."[All Fields] AND 2011[All Fields] no indexador PubMed do NCBI. Foram selecionadas manualmente quatro espécies de forma aleatória, e aquelas que possuíam na descrição os seguintes parâmetros:

- (1) o estudo filogenético com uso do 16S rRNA;
- (2) as espécies bacterianas recentemente descritas, deveriam ter o índice de similaridade/identidade;
- (3) o número de acesso das sequências mais similares obtidas através do alinhamento do BLAST e;
- (4) todas as sequências utilizadas no estudo depositadas no NCBI/DDBJ/EMBL.

As espécies submetidas ao fluxo de análise do BACSearch foram:

- *Chryseomicrobium imtechense* (GenBank Accession - GQ927308.1), descrita por Arora et al. (2011);
- *Desulfosoma caldarium* (GenBank Accession - FJ491989.1), descrita por Baena et al. (2011);

- *Calditerricola satsumensis* (GenBank Accession AB250968.1), descritas por Moriya et al. (2011); e
- *Calditerricola yamamurae* (GenBank Accession AB308475.1), descritas por Moriya et al. (2011).

De acordo com as descrições iniciais para cada nova bactéria, os autores registraram os resultados do alinhamento das bactérias que possuíam o maior índice de similaridade, tendo como base os resultados desse alinhamento do BLAST, utilizando a base do NCBI (Tabela 4.1).

Tabela 4.1: Espécie utilizada na validação do software BACSearch.

Espécies recentemente descritas	Espécie mais similar	Índice de Identidade (%)
<i>Chryseomicrobium imtechense</i>	<i>Psychrobacillus psychrodurans</i>	96,15%
<i>Desulfosoma caldarium</i>	<i>Desulfaginum hydrothermale</i>	93,00%
<i>Calditerricola satsumensis</i>	<i>Planilum yunnanense</i>	90,00%
<i>Calditerricola yamamurae</i>	<i>Planilum yunnanense</i>	90,00%

¹Fonte: (ARORA et al., 2011; BAENA et al., 2011; MORIYA et al., 2011)

4.2.2 MolEpi

Obtenção e processamento das amostras

O estudo foi realizado na Unidade de Emergência Dr. Daniel Houly ou Unidade de Emergência do Agreste (UEA), hospital referência no atendimento a casos de trauma na região do sertão, agreste e baixo São Francisco do estado de Alagoas, localizado no município de Arapiraca-Al (Figura 4.2).

Trata-se de um hospital público com 61 leitos distribuídos em 01 Unidade de Terapia Intensiva Geral, 08 Enfermarias para Internação e 05 Enfermarias para Emergência. Os locais selecionados para realização do estudo foram a UTI Geral e as Enfermarias de Internação.

Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Alagoas sob o número 23065.014852/2006-29 (Anexo 1), através da proposta Vigilância Epidemiológica na Unidade de Emergência do Agreste: Implantação de Serviço de Epidemiologia Molecular e Criação de Banco de Dados Informatizado para sequências de rDNA 16S de bactérias patogênicas, submetida a chamada PP-SUS / 2006 do Ministério da Saúde.



Figura 4.2: Mapa do estado de Alagoas.

Todos pacientes e/ou seus responsáveis, foram esclarecidos sobre o estudo e a coleta do material que somente foi realizada quando concordaram em participar do mesmo (Apêndice B).

As amostras coletadas na UEA foram analisadas no Laboratório de Biologia Molecular e Expressão Gênica – LABMEG, da Universidade Federal de Alagoas – UFAL, Campus Arapiraca. Durante um período de 24 meses (março de 2008 a março de 2010), um total de 200 bactérias foram selecionadas para análise. As amostras foram obtidas de várias origens: 124 exsudatos purulentos de feridas cirúrgicas, 30 exsudatos purulentos de escaras, 31 culturas de urina e 15 culturas de aspirado traqueal.

A coleta microbiológica foi realizada através de busca ativa dos casos de infecção, por observação direta dos pacientes e análise de prontuário, quando este evidenciava características sugestivas de quadros infecciosos ou mesmo a própria solicitação médica para exame microbiológico.

Extração de DNA

As colônias bacterianas obtidas de isolados clínicos foram submetidas ao processo de extração de DNA, seguindo o protocolo do *Current Protocols in Molecular Biology* (AUSUBEL et al., 2003). Algumas colônias bacterianas da mesma amostragem foram inoculadas em caldo de cultura BHI estéril. Após incubação em estufa a 35,1°C por 18h a 24h, a cultura foi transferida para micro-tubo de 1,5mL e centrifugada a 10.000 rpm por 2 min. O sobrenadante foi descartado e o *pellet* (centrifugado)

foi ressuspendido por inversão em 570 μ L de água *MilliQ*. Foram adicionados 30 μ L de SDS a 10% (*Sodium Dodecyl Sulfate*) e 3 μ L de proteinase *K* para lise da parede celular e digestão de proteínas. Após homogeneizar e incubar a reação por 1h a 37°C, adicionou-se 100 μ L de NaCl 5mM homogeneizando inteiramente.

Para extração utilizou-se o CTAB (Brometo de cetiltrimetilamônio), adicionando-se 80 μ L de CTAB NaCl, misturando-os por inversão, seguida de incubação a 65°C por 10 min. Polissacarídeos e proteínas residuais foram removidas em 2 extrações: com a adição de 800 μ L de clorofórmio / álcool isoamílico 24:1, sendo misturados e centrifugados a 10.000 rpm por 5 min; a fase aquosa é removida para um novo tubo e em seguida a outra extração com o uso de igual volume de fenol / clorofórmio / álcool isoamílico 25:24:1, também sendo misturados e centrifugados a 10.000 rpm por 5 min.

Após transferir o sobrenadante para um novo tubo (1,5mL), a precipitação dos ácidos nucleicos foi realizada adicionando-se 0,6 volume de isopropanol, misturando por inversão e centrifugação a 14.000 rpm por 5 min. Descartando-se o isopropanol por inversão do tubo, o *pellet* foi lavado com 1mL de etanol 70%, seguido de outra centrifugação a 14.000 rpm por 5 min. Posteriormente o etanol foi desprezado, retirando-se o excesso ao colocar o microtubo em banho seco a 60°C, sem deixar o *pellet* ressecar. Ao final, o *pellet* foi ressuspendido em 100 μ L de água *MilliQ* e a amostra estocada a -20°C.

PCR de amplo espectro e eletroforese

Para a reação de PCR de amplo espectro, foi selecionado o par de iniciadores 16Sf 5'-AGAGTTTGATCATGGCTCA-3' (iniciador direto, localizado nas posições 8 a 27) e 16Sr 5'-ACGGCGACTGCTGCTGGCAC-3' (iniciador reverso, posições 531 a 514) (GATSELIS et al., 2006). Estes iniciadores delimitam a região inicial do gene rRNA 16S de bactérias permitindo a amplificação parcial do gene. As sequências dos oligonucleotídeos foram submetidas a alinhamentos no programa BLAST (disponível em www.ncbi.nlm.nih.gov/BLAST), para confirmação de suas homologias com espécies bacterianas.

A reação de PCR para amplificação parcial do gene rRNA 16S foi realizada com os reagentes e volumes descritos na Tabela 4.2, com volume total de 25 μ L por reação.

Tabela 4.2: Reação da PCR para amplificação parcial do gene rRNA 16S.

Componentes	Volume na reação	Concentração final
H_2O ultrapura estéril	19,15 μL	-
Tampão 10x	2,5 μL	1x
$MgCl_2$ (50mM)	0,75 μL	1,5 mM
Mistura de dNTP (10mM)	0,5 μL	0,2 mM
Iniciador 16Sf (10 pmol/ μL)	0,5 μL	0,2 pmol/ μL
Iniciador 16Sr (10 pmol/ μL)	0,5 μL	0,2 pmol/ μL
Taq polimerase (5U/ μL)	0,1 μL	0,02 U/ μL
DNA (50ng/ μL)	1,0 μL	2 ng/ μL
Volume total	25 μL	

²Fonte: Autor, 2012.

Os ciclos de amplificação foram realizados em termociclador automático, modelo MJ96 (Biocycler) e consistiram de uma incubação inicial a 95°C por 5 min, seguida de 30 ciclos de desnaturação a 94°C por 30 segundos, anelamento a 55°C por 30 segundos e extensão a 72°C por 1,5 min, com uma extensão final a 72°C por 10 min. Os produtos de amplificação foram submetidos à eletroforese em gel de agarose 1,5% e visualizados sob luz ultravioleta. O tamanho do amplicon esperado foi de aproximadamente 500 pares de base.

Purificações e reação de sequenciamento

O procedimento utilizado na purificação da reação de PCR foi adicionado 72 μL de isopropanol 65% diretamente aos micro-tubos do amplicon, sendo homogeneizados e incubados 20 min a temperatura ambiente. As amostras foram centrifugadas a 14.000 rpm por 15 min, sendo o isopropanol descartado por inversão. Foi adicionado 250 μL de etanol 60% com nova centrifugação a 14.000 rpm por 5 min. Após descartar o sobrenadante e secar o *pellet*, este foi ressuscitado em 20 μL de água *MilliQ*. Protocolo adaptado do *Current Protocols in Molecular Biology* (AUSUBEL et al., 2003).

Uma alíquota da reação foi submetida à eletroforese em gel de agarose 2% para visualização e quantificação do DNA purificado. A reação de sequenciamento foi realizada em um volume final de 10 μL , contendo aproximadamente 10 μL do produto da PCR, 2 μL de Big Dye Mix (Applied Biosystems), 2 μL de Save Money 5.0X (Applied Biosystems) e 2 pmol de iniciador direto ou reverso (16Sf ou 16Sr). As amostras foram submetidas a 25 ciclos de desnaturação a 96°C por 10 segundos, anelamento a 50°C por 5 segundos e extensão a 60°C por 4 min, em termociclador

automático.

As amostras foram novamente purificadas, adicionando-se 72 μ L de isopropanol 65% sendo homogeneizadas e incubadas 20 min em temperatura ambiente, protegidas de luminosidade. A seguir foram centrifugadas a 14.000 rpm por 15 min, sendo o isopropanol descartado por inversão. Foi adicionado 250 μ L de etanol 60% com nova centrifugação a 14.000 rpm por 5 min. Após descartar o sobrenadante e secar o *pellet*, este foi ressuspendido em 20 μ L de Formamida HI-DI, sendo armazenado a -20°C, protegido da luminosidade. As amostras foram submetidas a sequenciamento automático nos equipamentos ABI PRISM 310 e ABI PRISM 3100 (Applied Biosystems).

Todos os produtos do sequenciamento, ou seja, as sequências do gene rRNA 16S, foram depositadas no GenBank (NCBI) e os número de acesso estão no Apêndice C.

5 Resultados e Discussão

5.1 Sistemas desenvolvidos

Esta seção descreve os sistemas desenvolvidos, que atenderam a demanda da presente proposta, o BACSearch e o MolEpi. O BACSearch caracteriza-se como um sistema de busca de sequências por homologia, integração otimizada de bancos de dados biológicos e programas de alinhamentos múltiplos de sequências (MSA) e estudos filogenéticos. O MolEpi é um sistema de gestão da informação do prontuário eletrônico do paciente, que possui o BACSearch integrado para identificação de bactérias patogênicas, o qual vincula essa identificação ao referido prontuário.

O MolEpi foi desenvolvido para atender como campo de aplicação a Unidade de Emergência do Agreste de Alagoas (UEA), pertencente ao Sistema Único de Saúde (SUS). Com a possibilidade de integração a uma rede regional ou nacional de diagnóstico molecular de microrganismos.

Os sistemas são caracterizados por serem de código aberto e livre, e foram registrados no Instituto Nacional de Propriedade Industrial - INPI como os seguintes protocolos: BACSearch: protocolo 000194; e o MolEpi: protocolo 000222 (Anexo 2).

5.1.1 BACSearch

O BACSearch é um programa baseado em um serviço de análise através da web, e foi desenvolvido para atender as seguintes características: (i) a integração de bancos de dados de referência em uma única aplicação; (ii) implementado fluxo de trabalho de recuperação de sequências de regiões gênicas utilizadas para estudo de identificação de bactérias; (iii) recuperação de sequências de acordo com os resultados de similaridade implementados através do BLAST, e a possibilidade desses resultados serem automaticamente incorporados em estudos filogenéticos através dos principais programas de alinhamento múltiplo de sequências, como CLUSTALW,

MUSCLE e MAFFT; (iv) proporciona à análise dos dados de forma a possibilitar estudos filogenéticos com os editores de alinhamento múltiplo, como JalView e o Archeopteryx; e (v) é um sistema de processamento otimizado, através de uma interface web fácil e ágil que proporciona à análise dos resultados de saída claros (Figura 5.1).

O fluxo de trabalho desenvolvido para o BACSearch pode ser visualizado na Figura 5.2. A identificação começa com o produto do sequenciamento, que em nosso estudo foram as sequências do gene rRNA 16S. O sistema utiliza a busca por similaridade através do BLAST, nos principais bancos de dados de referência (Figura 5.2 - A).

Para cada banco de dados de referência são gerados resultados que ficam disponíveis em abas (Figura 5.1 - A). A busca realizada através do BACSearch é integrada, e pode ser realizada em diversos bancos de dados (Figura 5.2 - B). O resultado do alinhamento do BLAST para cada sequência (*query*) pode ser realizado novamente, com outros parâmetros, e inserida para o andamento do estudo de alinhamento múltiplo de sequência (MSA) (Figura 5.2 - C).

A partir do resultado do alinhamento de sequências, é formado um arquivo de alinhamento múltiplo de sequência (MSA), que pode ser trabalhado com os programas ClustalW2, MAFFT ou MUSCLE (Figura 5.2 - D/E). As sequências semelhantes que obtiveram melhor similaridade são classificadas e acrescentadas ao arquivo em formato FASTA para alinhamento múltiplo. O resultado do alinhamento múltiplo fornecem três arquivos diferentes contendo, (i) o alinhamento da sequência de consulta ordenado por similaridade, (ii) uma árvore em formato Newick que é uma das maneiras de representação de árvores filogenéticas através de gráficos com comprimentos da borda usando parênteses e vírgulas, e (iii) as distâncias filogenéticas. O alinhamento obtido e as árvores filogenéticas são visualizados pelo Jalview (Figura 5.2 - F) e/ou pelo Archeopteryx. Assim, a identificação bacteriana e o posicionamento da sequência de consulta pode ser realizada através de observações minuciosas das árvores gerada por esses sistemas.

BacSearch: Web Server

BacSearch: an integrated webserver open-source for identifying optimal high-performance of bacteria by genomic sequences

Home About Tutorial Team

Blast submitted to: GREENGENE

Blast Data **GREENGENE** **MSA Data**

Data

Sequence:

```
GGAACCTCAAAGGAGATCCAGAGTTTGACATGGCTCAGATTGAAC
GCTGGCGGCAGGCCAACACATGCAAGTCGAGCGGTAACAGAAG
AAAGCTTGCTTTCTTGCTGACGAGCGCGGACGGGTGAGTAATG
TATGGGGATCTGCCCGATAGAGGGGGATACTAGGAAACGGT
GGCTAATACCGCATAATGTCCTCCGACCAAGCAGGGGATCTTC
GGACCTCGCACCACCGATGAGCCCATATGGGATTAGCTAGTAG
GTGGGGTAAAGGCTCACCTAGGCGACGATCTCTAGCTGGTCTGA
GAGGATGATCAGCCACACTGGAACCTGAGACACGGTCCAGACTCC
TACGGGAGGCAGCAGTGGGGAATTGCACAAATGGGCGCAAGCC
```

Database:

☐ GENBANK ☒ GREENGENE ☐ RDP2 ☐ SILVA

Submit

Blast Data **GREENGENE** **MSA Data**

Information

Database: GREENGENE

Query sequence length: 1142

Description

ID	Description	Score	E-value	MSA
gnl BL_ORD_ID 214925	281826 EU643833.1 Proteus mirabilis str. Hu	1029.0	0.0	+ MSA (all)
gnl BL_ORD_ID 22042	27167 AF008582.1 Proteus mirabilis str. ATCC 29906T	1021.0	0.0	+ MSA
gnl BL_ORD_ID 43355	108937 AY820623.1 Proteus mirabilis	1017.0	0.0	+ MSA
gnl BL_ORD_ID 96799	161792 DQ885256.1 Proteus mirabilis str. NCTC 11938	1011.0	0.0	+ MSA
gnl BL_ORD_ID 20256	19524 AJ301682.1 Proteus mirabilis str. CIP103181T (ATCC29906T)	1011.0	0.0	+ MSA

Blast Data **GREENGENE** **MSA Data**

Data

Program: Clustalw

Skip user sequence: No

ID	Database	Description	Score	E-value	MSA
gnl BL_ORD_ID 214925	GREENGENE	281826 EU643833.1 Proteus mirabilis str. Hu	1029.0	0.0	- MSA (all)
gnl BL_ORD_ID 22042	GREENGENE	27167 AF008582.1 Proteus mirabilis str. ATCC 29906T	1021.0	0.0	- MSA
gnl BL_ORD_ID 43355	GREENGENE	108937 AY820623.1 Proteus mirabilis	1017.0	0.0	- MSA
gnl BL_ORD_ID 96799	GREENGENE	161792 DQ885256.1 Proteus mirabilis str. NCTC 11938	1011.0	0.0	- MSA
gnl BL_ORD_ID 20256	GREENGENE	19524 AJ301682.1 Proteus mirabilis str. CIP103181T (ATCC29906T)	1011.0	0.0	- MSA

Submit

Blast Data **GREENGENE** **MSA Data** **CLUSTALW**

Information

Program: CLUSTALW

Description

Visualization

JalView: can be used to perform the MSA Analysis (requires Java) **Start JalView**

Download

Alignment File: can be used with JalView and several other programs [.aln](#)

Dendogram File: can be used with TreeView and several other programs [.dnd](#)

Input File: fasta file with the sequences chosen for analysis [.infile](#)

©2010 BacSearch

Figura 5.1: Interface gráfica do software BACSearch. A - interface para inserção de sequências; B - resultado do alinhamento do BLAST; C - lista dos resultados da busca por similaridade, demonstrando *score* e o *e-value*; D - agrupamento de resultados para alinhamento múltiplo; e E - resultado do alinhamento múltiplo de sequências para visualização dos editores de alinhamento e estudo filogenético.

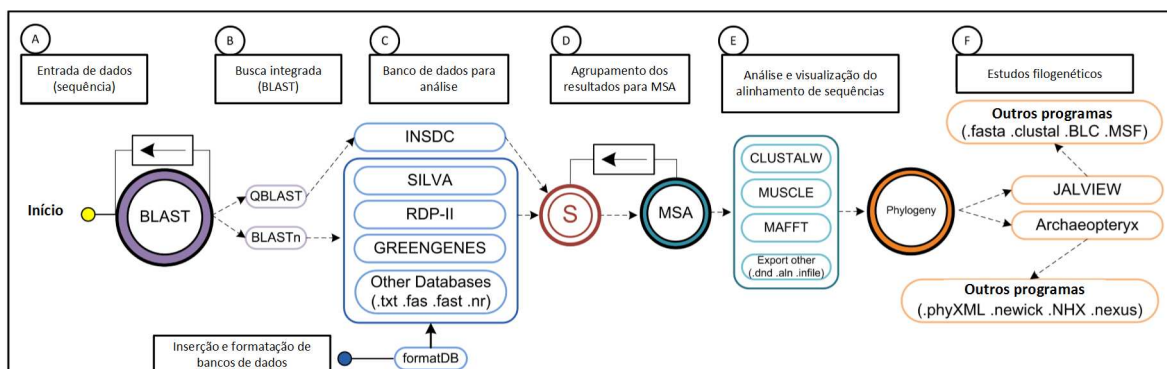


Figura 5.2: Fluxo de trabalho do BACSearch. A - Entrada da sequência; B - Alinhamento integrado de sequência através do BLAST em bancos de dados de referência; C - Bancos de dados de referência; D - O resultado do alinhamento do BLAST são inseridos para o estudo de MSA; E - Programas utilizados no MSA; F - Os resultados do MSA são utilizados em estudos de árvores filogenéticas, com Jalview e/ou Archeopteryx.

5.1.2 MolEpi

Modelagem do sistema do banco de dados do MolEpi

Todas as informações foram modeladas em um diagrama de entidade-relacionamento (ER), o qual possibilitou a implementação do sistema (Figura 5.3).

A modelagem e implementação do banco de dados para o gene rRNA 16S do MolEpi foi realizada de forma que todos os dados possuíssem entidades e atributos suficientemente necessários para que o mapeamento das relações existentes entre os dados dos pacientes e as informações microbiológicas e laboratoriais fossem armazenadas de forma lógica e posteriormente recuperadas e comparadas eficientemente.

O banco de dados do MolEpi foi desenvolvido utilizando o modelo centralizado no prontuário do paciente. Assim, o modelo expressa um relacionamento inicial entre entidades, que no banco de dados estar representado como um tabela de dados relacional (Figura 5.3), possuindo cinco áreas de domínios inter-relacionadas:

1. Informações do paciente (Figura 5.3 - A);
2. Informações da amostra coletada (Figura 5.3 - B);
3. Resultados dos alinhamentos de sequências (Figura 5.3 - C)
4. Teste laboratorial de resistência e microbiológico (Figura 5.3 - D);
5. Identificação taxonômica (Figura 5.3 - E).

O sistema MolEpi foi desenvolvido para ser utilizado em etapas de inserção dos dados e consultas como observado na Figura 5.4. O sistema é composto por cinco interfaces com o usuário que estão envolvidos na alimentação e na análise desses dados. Essas etapas são subdivididas da seguinte forma:

Informação do Paciente: nessa interface são obtidas todas as informações do paciente necessárias para iniciar o processo de coleta das amostras; Figura 5.4 - B;

Informação da Amostra: nesse momento (Figura 5.4 - C) são inseridas as informações sobre as amostras que foram coletadas;

Microbiologia: no laboratório de microbiologia são realizadas a cultura, isolamento e testes de resistência, todas essas informações são armazenadas no sistema Figura 5.4 - D;

Análise Molecular: após o isolamento em cultural inicial, a amostra é encaminhada para extração, purificação e amplificação para posterior sequenciamento, e o produto do sequenciamento, é inserido no sistema; Figura 5.4 - E;

Bioinformática: a identificação e análise do produto do sequenciamento, ou seja, a sequência do gene rRNA 16S, é identificada através de ferramentas utilizando o *workflow* do BACSearch, que encontra-se integrado ao MolEpi; Figura 5.4 - F;

Todas essas interfaces de comunicação possuem como foco principal o prontuário eletrônico do paciente (Figura 5.4 - G), e a recuperação dessas informações que estão contidas no MolEpiDB (Figura 5.4 - H) são realizadas por essa interface. As demais interfaces do programa estão no Apêndice E.

Para a identificação e análise dos microrganismos patogênicos foram desenvolvidos dois mecanismos de estudo. O primeiro foi a identificação através do alinhamento de sequências em bancos de dados de referência para o gene rRNA 16S. Nessa fase, a integração do BLAST (Figura 5.4 - I) como ferramenta de recuperação e identificação de sequências através de similaridade é realizada em bancos de dados de referência (Figura 5.4 - L). O outro mecanismo foi o estudo das sequências mais similares através da interface de análise filogenética (Figura 5.4 - M), que uti-

lizam sistemas de recuperação de sequências através das informações armazenadas no registro eletrônico do paciente (Figura 5.4 - N).

5.2 Validação do BACSearch

De acordo com dados das novas descrições de espécies, foi realizado a validação do *workflow* proposto para o BACSearch. As espécies submetidas para a estudo no BACSearch foram *Chryseomicrobium imtechense* descrita por Arora et al. (2011), *Desulfosoma caldarium* descrita por Baena et al. (2011), *Calditerricola satsumensis* e *Calditerricola yamamurae* descritas por Moriya et al. (2011), mais detalhes na Tabela 4.1.

As sequências das novas espécies descritas na Tabela 4.1, foram analisadas, e os resultados estão listados na Tabela 5.1.

A nova espécie *Chryseomicrobium imtechense* retornou como mais idêntica as espécies *Planococcaceae bacterium* (98,95%) no GenBank (NCBI) e no RDP-II (Tabela 5.1 - linha 1.1 e 1.2), *Crocino bacterium jeju* (98,95%) SSURefTax(ARB) (Tabela 5.1 linha 1.4) e *Shivajiella amylolytica* (98,95%) ProkMSA(LBL) (Tabela 5.1 - linha 1.6), com uma diferença de 2,80 pontos percentuais positivos com relação as espécies citadas como mais idênticas segundo Arora et al. (2011). Essa diferença entre a espécie mais idêntica citada na descrição original e os resultados do BACSearch demonstram que *Chryseomicrobium imtechense* possui mais identidade com as espécies do Filo Firmicutes, Classe Bacilli, Ordem Bacillales, e não com *Perluclidibaca piscinae* que possui posicionamento taxonômico no filo Proteobacteria, classe Gammaproteobacteria, ordem Pseudomonadales.

Desulfosoma caldarium obteve como mais idêntica a espécie *Desulfatimicrobium mahresensis* com 99,39% de identidade, com uma variação de 6,39 pontos percentuais positivos com relação à *Desulfacinum hydrothermale* (Tabela 5.1 - linha 2.1 - 2.4 e 2.6). Embora a variação encontrada seja considerável, o posicionamento taxonômico não foi alterado permanecendo no mesmo gênero.

As espécies *Calditerricola satsumensis* e *Calditerricola yamamurae* foram descritas por Moriya et al. (2011) e obtiveram índices de identidade significativamente diferenciados com relação aos descritos pelos autores (Tabela 5.1). Quando analisados os resultados do alinhamento para a espécie *Calditerricola satsumensis*, o

BACSearch retornou como maior índice de identidade a espécie *Caldiclava thermospodii* com 99,17%, uma variação de 9,17 pontos percentuais em comparação com os resultados apresentados pelos autores Moriya et al. (2011) (Tabela 5.1 - linha 3.3 e 3.6). Para a espécie *Calditerricola yamamurae* o alinhamento forneceu como mais idêntica a espécie *Caldiclava thermospodii* com 98,62% de similaridade, com os bancos de dados Greengenes (LBL) e ProkMSA (LBL) (Tabela 5.1 - linha 4.3 e 4.6). Nesses dois casos o posicionamento da espécie mais similar citado por Moriya et al. (2011), foi *Planifilum yunnanense*, que não pertence ao mesmo gênero de *Calditerricola*, a qual pertence a família Thermoactinomycetaceae, sendo *Planifilum yunnanense*, a família Bacillaceae.

Esses resultados demonstram que as análises de identificação de novas bactérias com uso do gene rRNA 16S, quando realizadas em um único banco de dados, retornam resultados que podem demonstrar a verdadeira relação ou posicionamento filogenético esperado. Quando esse estudo é realizado de forma integrada, como os proporcionadas pelo BACSearch, os resultados tornam-se mais conclusivos, retornando valores de similaridade mais próximas das relações evolutivas verdadeiras.

As variações de um posicionamento filogenético ou uma identificação imprecisa de bactérias através do gene rRNA 16S, em áreas de aplicação como a clínica, podem levar a uma terapêutica vaga, principalmente quando considera-se a evolução clínica. Observando também que a identificação de bactérias, principalmente aquelas espécies de difícil cultivo ou novas, que são raramente isoladas através de protocolos de laboratórios clínicos ou cepas bioquimicamente aberrantes, somente podem ser discriminadas através de análises moleculares por sequenciamento (PATEL, 2001; CLARRIDGE, 2004).

Podemos considerar que um dos grandes impedimentos no uso do gene rRNA 16S, é que mesmo existindo um grande número de pacotes de softwares e bancos de dados específicos para identificação de bactérias, os resultados proporcionados por estes não são claros e muitas vezes de difícil compreensão, quando analisados por usuários inexperientes e com o uso de somente uma base de dados. BACSearch pode ser usado para o ensino em bioinformática, como busca em bancos de dados, alinhamento múltiplo de sequências e estudo filogenéticos por proporcionar uma interface amigável e fácil de usar.

Tabela 5.1: Identificação obtida através do software BACSearch.

Query (Spp.)	Subject	Dif.(pts.)	Ident.(%)	Databases 16S
1.1	<i>Planococcaceae bacterium</i>	2,80	98,95	GenBank (NCBI)
1.2	<i>Planococcaceae bacterium</i>	2,80	98,95	RDP-II(CME/MSU)
1.3	<i>Planococcus citreus</i>	0,13	95,88	Greengenes (LBL)
1.4	<i>Crocino bacterium jejui</i>	2,80	98,95	SSURefTax(ARB)
1.5	<i>Bacillus sphaericus</i>	2,09	94,06	GoldStrains (LBL)
1.6	<i>Shivajiella amylolytica</i>	2,80	98,95	ProkMSA(LBL)
1.7	<i>Bacillus psychrodurans</i>	0,29	96,44	RefSeq(NCBI)
2.1	<i>Desulfatimicrobium mahresensis</i>	6,36	99,36	GenBank (NCBI)
2.2	<i>Desulfatimicrobium mahresensis</i>	6,36	99,36	RDP-II(CME/MSU)
2.3	<i>Desulfatimicrobium mahresensis</i>	6,36	99,36	Greengenes (LBL)
2.4	<i>Desulfatimicrobium mahresensis</i>	6,36	99,36	SSURefTax(ARB)
2.5	<i>Desulfoglaeba alkanexedens</i>	0,17	93,17	GoldStrains (LBL)
2.6	<i>Desulfatimicrobium mahresensis</i>	6,39	99,39	ProkMSA(LBL)
2.7	<i>Desulfacinum hydrothermale</i>	0,59	93,59	RefSeq(NCBI)
3.1	<i>Planifilum yunnanense</i>	0,41	90,41	GenBank (NCBI)
3.2	<i>Geobacillus jurassicus</i>	1,00	91,00	RDP-II(CME/MSU)
3.3	<i>Caldiclava thermospodii</i>	9,17	99,17	Greengenes (LBL)
3.4	<i>Geobacillus jurassicus</i>	1,00	91,00	SSURefTax(ARB)
3.5	<i>Geobacillus kaustophilus</i>	1,14	91,14	GoldStrains (LBL)
3.6	<i>Caldiclava thermospodii</i>	9,17	99,17	ProkMSA(LBL)
3.7	<i>Planifilum yunnanense</i>	1,66	91,66	RefSeq(NCBI)
4.1	<i>Planifilum yunnanense</i>	0,96	90,96	GenBank (NCBI)
4.2	<i>Geobacillus jurassicus</i>	1,60	91,60	RDP-II(CME/MSU)
4.3	<i>Caldiclava thermospodii</i>	8,62	98,62	Greengenes (LBL)
4.4	<i>Geobacillus jurassicus</i>	1,60	91,60	SSURefTax (ARB)
4.5	<i>Geobacillus kaustophilus</i>	1,68	91,68	GoldStrains (LBL)
4.6	<i>Caldiclava thermospodii</i>	8,62	98,62	ProkMSA(LBL)
4.7	<i>Planifilum yunnanense</i>	1,98	91,98	RefSeq(NCBI)

¹Nova espécie(Spp.); Diferença em pontos (Dif. pts.); Porcentagem de identidade(Ident.(%)).

²Fonte: Autor, 2012.

5.3 Validação do MolEpi

Duzentas (200) bactérias tiveram o gene RNA ribossomal 16S extraído e posteriormente amplificado, resultando em *amplicons* de aproximadamente 500 pares de base. Em seguida foram submetidos à reação de sequenciamento, com sequências de 446 ± 54 nucleotídeos em média, e foram depositadas no GenBank (NCBI) (Ver Apêndice C).

Com o sistema integrado de busca em bancos de dados específicos de 16S rRNA foi possível identificar os 200 casos em nível de espécie (Tabela 5.2). Os resultados de todos os alinhamentos das sequências do 16S rRNA estão no Apêndice D.

As bactérias mais frequentes em nosso estudo foram do gênero *Pseudomonas* com 48 (24%) sendo isolados 46 de *P. aeruginosa*; 39 (19,5%) de *Acinetobacter* sendo

que 38 casos eram de *A. baumannii*, e 33 (16,5%) para o gênero *Klebsiella*, sendo 31 (15,5%) para *K. pneumoniae* e 27 (13,5%) *Staphylococcus* sendo 18 (9%) *S. aureus*.

Tabela 5.2: Espécies identificadas através do MolEpi.

Qtd.	Gênero	Qtd.	Espécie
46	<i>Pseudomonas</i>	42	<i>P. aeruginosa</i>
		02	<i>P. cedrina subsp. fulgida</i>
		02	<i>P. putida</i>
39	<i>Acinetobacter</i>	38	<i>A. baumannii</i>
		01	<i>A. radioresistens</i>
33	<i>Klebsiella</i>	31	<i>K. pneumoniae</i>
		02	<i>K. granulomatis</i>
27	<i>Staphylococcus</i>	18	<i>S. aureus</i>
		04	<i>S. epidermidis</i>
		02	<i>S. haemolyticus</i>
		02	<i>S. hominis</i>
		01	<i>S. saprophyticus</i>
		01	<i>S. sciuri</i>
		01	<i>E. cloacae</i>
		14	<i>E. hormaechei</i>
12	<i>Escherichia</i>	12	<i>E. coli</i>
05	<i>Enterococcus</i>	04	<i>E. faecalis</i>
		01	<i>E. gallinarum</i>
04	<i>Proteus</i>	04	<i>P. mirabilis</i>
04	<i>Corynebacterium</i>	04	<i>C. striatum</i>
04	<i>Serratia</i>	04	<i>S. marcescens</i>
03	<i>Morganella</i>	03	<i>M. morganii subsp morganii</i>
02	<i>Providencia</i>	02	<i>P. stuartii</i>
01	<i>Bacillus</i>	01	<i>B. clausii</i>
01	<i>Brevibacillus</i>	01	<i>B. agri</i>
01	<i>Stenotrophomonas</i>	01	<i>S. maltophilia</i>
Total		200	

³Fonte: Autor, 2012.

Os índices de identificação das bactérias patogênicas em espécies através do alinhamento de sequências através do MolEpi variaram entre 96,45% a 100,00%. Os altos índices de similaridades obtidos com a utilização de diversos bancos de dados, tornaram os resultados mais consistentes, possibilitando à identificação para 100% das espécies estudadas. Em sua grande maioria, os índices da análise por alinhamento de sequências obtidas com a integração dos bancos de dados específicos para o gene rRNA 16S exibem maior percentagem de identificação em espécie do que os métodos convencionais e comerciais, e do que outros publicados na literatura (WOO et al., 2008). Estudos realizados por Bosshard et al. (2003), Bosshard et al. (2004), Mignard e Flandrois (2006), Woo et al. (2008) e demais colaboradores demonstram que o sucesso na identificação através do alinhamento tem variado de 62% a 83% para o táxon espécie, levando-se em consideração qual grupo e os critérios

de obtenção das amostras. Com a integração dos bancos de dados de gene rRNA 16S obtivemos 200 identificações conclusivas em espécies, como podemos observar no Apêndice D.

Embora o MolEpi tenha proporcionado uma elevada taxa de sucesso nos alinhamentos, existem grandes dificuldades na identificação de bactérias, principalmente para os gêneros *Burkholderia* e *Yersinia*, no qual a análise por alinhamento de sequências apresenta-se não tão conclusivo (WOO et al., 2008). Entretanto, grande parte das identificações realizadas em laboratórios de microbiologia possui como foco principal bactérias patogênicas, e em alguns casos os métodos fenotípicos não são conclusivos para algumas cepas. O uso de métodos de identificação por alinhamento de sequência tem levado a um reconhecimento epidemiológico e patogênico mais aprofundado dessas bactérias (CLARRIDGE, 2004).

Temos que considerar para a identificação de bactérias patogênicas as regiões mais conclusivas para sequenciamento e posterior identificação por alinhamento. Os estudos realizados por Chakravorty et al. (2007) demonstraram que as regiões hipervariáveis de nucleotídeos V1 (69-99), V2 (137-242) e V3 (433-497) presentes no gene rRNA 16S possibilita maior diferenciação entre os gêneros de *Staphylococcus* se considerarmos a região V1. Já região V2 é mais conclusiva na diferenciação em gênero de grande parte das bactérias patogênicas, excluindo as espécies de *Escherichia*, *Shigella*, *K. pneumoniae* e *Enterobacter aerogenes*. A principal característica da região V3, é a sua aplicação para a diferenciação das espécies da família Enterobacteriaceae, como as espécies do gênero *Enterobacter* e *Klebsiella*. Considerando os resultados dos alinhamentos do MolEpi, a identificação em gênero/espécie foi realizada de forma satisfatória para todos os casos.

As bactérias apresentadas na Tabela 5.2, em sua grande maioria, são agentes infecciosos de grande interesse em ambientes nosocomiais. Mesmo assim, obteve-se ocorrências de (*Corynebacterium amycolauum* e *Brevibacillus agri*), que são pouco descritos na literatura como bactérias de ambiente hospitalar e/ou raramente identificados através de métodos tradicionais, pois são difíceis de identificar ou por considerá-los uma contaminação amostral.

Foi identificado *Bacillus clausii*, que compreende um grande grupo de bacilos Gram positivos aeróbicos e facultativos que se caracterizam pela capacidade de formar esporos em condições aeróbicas, são utilizados na prevenção de infecções

respiratória recorrentes em crianças (DUC et al., 2004; MARSEGLIA et al., 2007; SINGHI; BARANWAL, 2008).

Considerando ainda microrganismos raros em diagnósticos clínicos, foi identificado através do alinhamento *Corynebacterium amycolatum*, um grande agente infeccioso que acomete paciente imunocomprometidos (FUNKE et al., 1997; LOPEZ; BAPTESTE, 2009; WONG et al., 2010; KIMURA et al., 2011). Ainda, Esteban et al. (1999) realizaram re-análises de isolados de *C. xerosis* no período de 1992 a 1997, e observaram que, tratavam-se das espécies *C. striatum* e *C. amycolatum*, que foi registrada em infecções por cateter ou feridas cirúrgicas, erroneamente identificadas. Ainda segundo os estudos realizados por Camello e al. (2003) e colaboradores, que demonstraram uma grande dificuldade na identificação de espécies através de métodos convencionais de *Corynebacterium* não-diftéricos, em pacientes de um hospital universitário do Rio de Janeiro – Brasil, chegaram a conclusão também que *C. amycolatum* pode ocorrer em qualquer sítio infeccioso, pois foi encontrado em todos os tipos de amostras analisadas.

Foi identificado o microrganismo *C. striatum* como causador de infecção do trato urogenital. *C. striatum* é encontrado normalmente como parte da microbiota das narinas do gado e na pele humana, e associado a uma variedade de infecções, principalmente em indivíduos imunocomprometidos. Entretanto, Lopez e Bapteste (2009) relataram um caso de infecção do trato urinário causada por *C. striatum* em uma mulher com cistite e hemácias na urina.

Um isolado de *Acinetobacter radioresistens*, foi detectado a partir de uma amostra de ferida, em um paciente com infecção pós-traumática. Os estudo referentes ao *A. radioresistens* caracterizam-no como um cocobacilo Gram negativo não-fermentador de ocorrência em ambientes naturais da pele humana (SEIFERT et al., 1997). Possui 4 a 8 vezes mais resistência à radiação gamma que os outros microrganismos do mesmo gênero (NISHIMURA et al., 1994; SAHA; CHOPADE, 2009). Estudos relatam bacteremia por *A. radioresistens* em pacientes de hospitais americanos detectados pelo projeto SCOPE - *Surveillance and Control of Pathogens of Epidemiologic Importance* (WISPLINGHOFF et al., 2000).

A espécie *Brevibacillus agri*, caracterizada como patógeno de difícil isolamento, foi identificada a partir de uma amostra de um paciente internado na enfermaria do hospital. Essa espécie pertencia ao gênero *Bacillus*, e foi reavaliado através

do sequenciamento e posterior análise de sequência do gene rRNA 16S, e proposto por Shida et al. (1996) e colaboradores, como novo gênero. Essa espécie é caracterizado por ocorrer em larvas de abelhas mortas, águas contaminadas e em fábricas de processamento de gelatina (LOGAN et al., 2002). O registro da Coleção de Cultura ARS (Agricultural Research Service) do Departamento de Agricultura dos Estados Unidos, cepa NRRL B-1157 refere ao único isolamento do microrganismo a partir de uma amostra clínica, em um paciente com febre ondulante.

Um dos isolados de *Pseudomonas* exibiu elevada similaridade com a espécie *P. cedrina subsp. fulgida* (Tabela 5.2). A amostra foi proveniente de um paciente vítima de perfuração por arma de fogo. Os estudos iniciais para essa espécie são de Dabboussi et al. (1999), os quais descreveram a espécie como *Pseudomonas cedrella* classificada como pertencente ao grupo I, chamado de grupo fluorescente. Nos estudos realizados por Behrendt et al. (2009), foi denominada *P. cedrina* devido as análises do gene rRNA 16S. Essa análise separaram em duas subespécies *P. cedrina subsp. fulgida* e *P. cedrina subsp. cedrina*.

Em uma amostra de um paciente vítima de acidente automobilístico, foi isolado o microrganismo *Pseudomonas putida* que é de ocorrência em águas, no solo e em reservatórios de água no ambiente hospitalar. Pode ocorrer normalmente na microbiota normal da faringe, e é patógeno oportunista em seres humanos. Já foi descrito como causador de diversas infecções, como do trato urinário (SAHA; JAIN; KAUR, 2010), bacteremia (YOSHINO et al., 2011), peritonite (DERVISOGLU et al., 2008), artrite séptica (MACFARLANE; OPPENHEIM; LORRIGAN, 1991), infecções em pacientes com câncer (ROLSTON et al., 2005), inclusive com detecção de genes que conferem resistências importantes, como as metallo-beta-lactamases (TREVINHO et al., 2010).

Na Unidade de Emergência do Agreste (UEA), foi isolado o microrganismo *Staphylococcus sciuri*. Inicialmente descrito por KLOOS, SCHLEIFER e SMITH (1976), é considerada uma espécie ancestral e um dos mais abundantes do gênero e caracterizado por ser de extensa distribuição na natureza. É caracterizado como comensal de espécies de roedores, marsupiais, e possuir registros em animais domésticos e animais de fazenda (HAUSCHILD; SCHWARZ, 2003). Os estudos realizados por Wallet et al. (2000) e Stepanovic et al. (2003) indicaram o crescimento da ocorrência de *S. sciuri* como agente patogênico, pois existem registro de infec-

ções causadas por esse microrganismos em endocardites, infecção do trato urinário e infecções de feridas.

Nas identificações realizadas através do alinhamento de sequências, obtêm-se a identificação dos microrganismos com o uso de bancos de dados biológicos e ferramentas como o BLAST, e em alguns casos não é possível determinar o nível taxonômico em gênero ou espécie. As identificações realizadas através de ferramentas de recuperação de sequências como BLAST, obtidos somente no GenBank/NCBI, forneceram alto índices de similaridade com diversas *subject*, não permitindo a diferenciação taxonômica. Com o uso dos bancos de dados de referência (RDP-II, SILVA e Greengenes) para o gene s rRNA 16S, foi obtido resultados mais consistentes.

Estudos realizados por Clarridge (2004) apresentam algumas teorias que explicam os resultados das inconsistências apresentadas pelo GenBank/NCBI. Os bancos de dados não-curados, possibilitam o depósito de qualquer sequência genética, permitindo que sequências com erros de anotação genômica de microrganismos possam ser depositadas. Esse erro leva a outro problema, o da não validação das sequências e informações relacionadas as mesmas, que são encaminhadas aos bancos de dados, pois estes aceitam qualquer nomenclatura dada às sequências depositadas (CLARRIDGE, 2004; ASHELFORD et al., 2005; JANDA; ABBOTT, 2007).

Existem também dificuldades relacionadas às características das bactérias, pois existem espécies que apresentam algumas regiões idênticas quando estudadas através do gene rRNA 16S, levando a altos índices de similaridade através do BLAST (CLARRIDGE, 2004). Além disso, quanto menor o tamanho do fragmento rDNA obtido através do processo de extração e sequenciamento, existe a necessidade de ser realizado nova extração para que não ocorra alinhamentos inconsistentes, bem como para alguns espécies a necessidade de ser realizado a obtenção de sequências para outras regiões variáveis, ou até mesmo o alinhamento de todo o gene rRNA 16S para se chegar a identificação em gênero/espécie (CHAKRAVORTY et al., 2007).

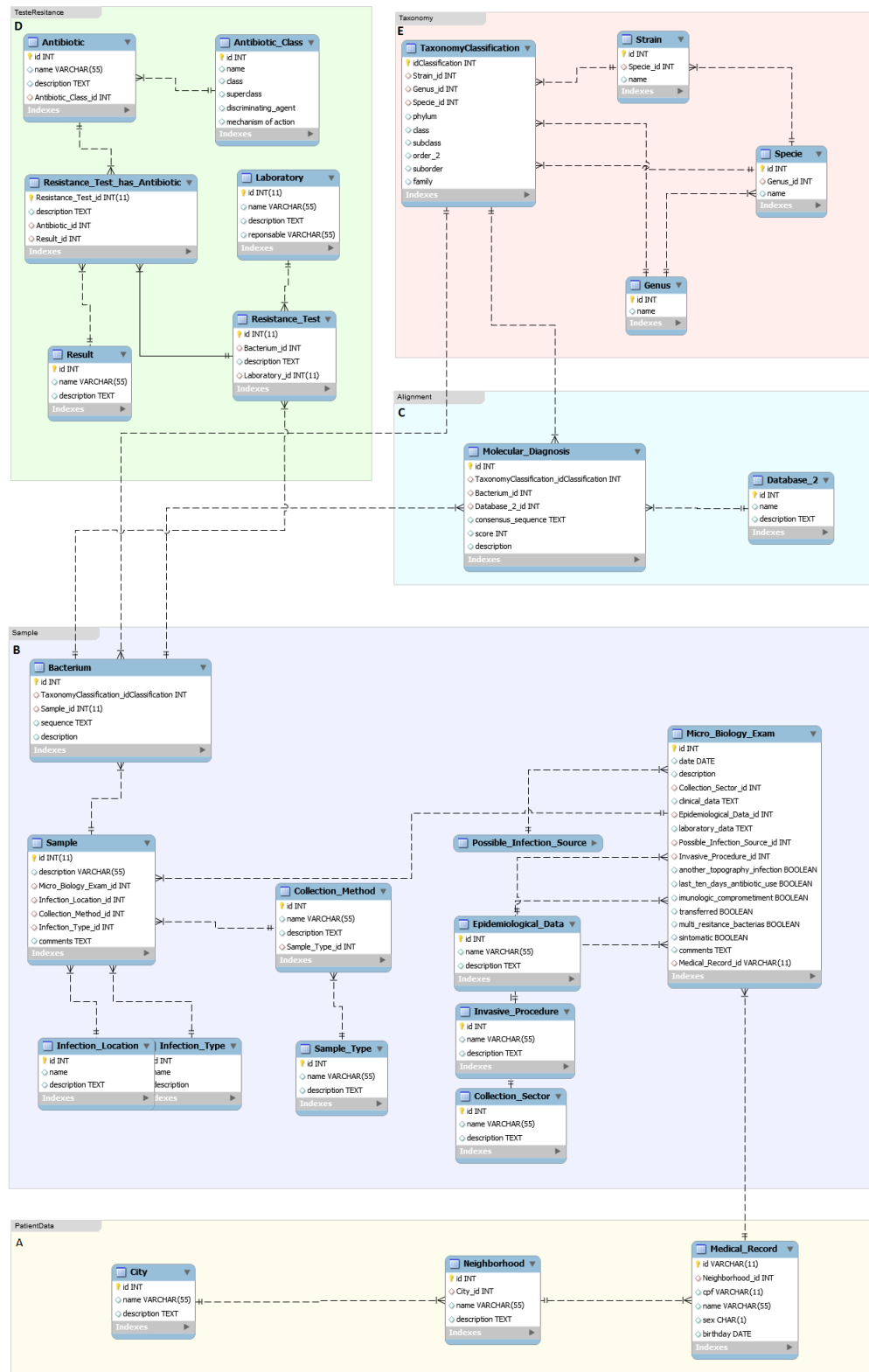


Figura 5.3: Modelo conceitual Entidade-Relacional do banco de dados MolEpi. A - Dados pessoais do paciente para iniciar o processo de análise; B - Nesse módulo do banco de dados que apresenta todas as tabelas que irão armazenar as informações das amostra obtidas e estão relacionadas com o registro eletrônico do paciente; C – Os registros dos alinhamentos obtidos são armazenados com o melhor resultado e o nome do banco de dados biológico; D – Todos os dados relacionados com a análise da amostra e que são realizados no laboratório de microbiologia são armazenados nesse conjunto de tabelas; E – A identificação realizada até o táxon mais específico obtido através dos alinhamentos são armazenados nesse conjunto de tabelas.

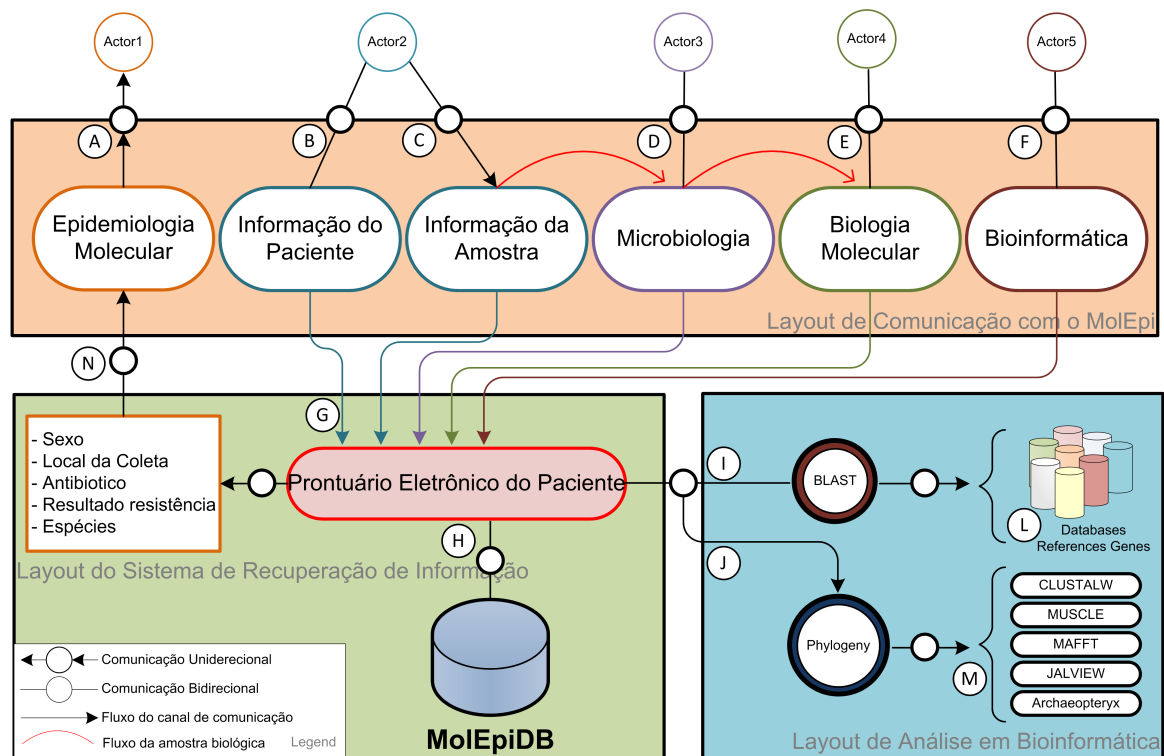


Figura 5.4: Fluxo de trabalho do MolEpi. A - Estudo das relações epidemiológicas no ambiente hospitalar; B - Informações do paciente; C - Inserção de informação das amostras biológicas; D - Estudo microbiológico da amostra em laboratório; E - Inserção do produto do sequenciamento; F - Identificação e análise do produto do sequenciamento; G - Informações centradas no Prontuário Eletrônico do Paciente; H - Recuperação de informações para análise das sequências; I - Identificação através do BLAST; J - Integração do BLAST como bancos de dados de referência, em L; M - Estudo das sequências através da interface de análise filogenética, que utilizam as informações armazenadas no registro eletrônico do paciente em N.

6 Conclusões

6.1 Considerações finais

Neste trabalho foram desenvolvidas duas ferramentas baseadas em técnicas de sistemas de informação, gestão da informação e de bioinformática. O MolEpi é um sistema de gerenciamento de informações que possui um banco de dados para sequências do gene rRNA 16S, visando o armazenamento que proporciona a recuperação de informações do paciente relacionadas, também com os resultados dos testes de resistência aos antimicrobianos. É um software de análise filogenética e interpretação de dados direcionados para a identificação de bactérias patogênicas, o BACSearch,

Com a incorporação no sistema de ferramentas de alinhamento, análise de dados moleculares e busca em bases de dados específicas para a região gênica do rRNA 16S, incluindo as bases de dados do NCBI (GenBank), RDP-II, SILVA e GreenGenes, e a possibilidade de inserir outras bases relevantes para o estudo, foi possível a identificação de microrganismos por sequenciamento do gene rRNA 16S.

Além disso, a análise integrada proporcionada pelo sistema demonstrou uma alta taxa de identificação, chegando a 100% de sucesso na determinação em espécie, inclusive de espécies raramente diagnosticadas através de metodologias fenotípicas. Portanto, configura-se um importante instrumento para elucidação de casos em que o diagnóstico convencional não permite tamanha precisão.

A adequada identificação de um microrganismo através da análise do gene rRNA 16S, por meio de iniciadores específicos que contemplem essa região, proporcionou a identificação de espécimes bacterianas raras, emergentes ou que não são rotineiramente identificados em laboratórios clínicos, como os microrganismos: *A. radioresistens*, *B. claussi*, *B. agri*, *C. amycolatum*, *C. striatum*, *P. cedrina* subsp. *fulgida*, *P. putida* e *S. sciuri*.

Assim, a utilidade dos sistemas desenvolvidos ficaram evidenciadas como seguem:

- Com relação ao BacSearch, foi possível realizar identificação de microrganismos através de sequências do rRNA 16S de forma otimizada, devido a flexibilidade e integração com os principais bancos de dados de rRNA 16S;
- Além disso, o sistema possui ferramentas de alinhamento múltiplo e análise filogenética, tornando-o mais completo, pois além de integrar os bancos de dados citados nesta pesquisa, todos os resultados podem ser analisados com ferramentas de bioinformática já validadas e consolidadas;
- O BACSearch também pode ser usado para o ensino em bioinformática, como busca em bancos de dados, alinhamento múltiplo de sequencias e estudo filogenéticos por proporcionar uma interface amigável e fácil de usar;
- O desenvolvimento do MolEpi, como sistema de prontuário eletrônico do paciente, integrado com o BacSearch, permitiu que todas as informações clínicas e epidemiológicas pudessem ser analisadas com mais precisão;
- O mapeamento das infecções dentro do ambiente hospitalar e a possibilidade de se realizar estudos com os resultados dos teste de resistências e os microrganismos patogênicos, faz do MolEpi um sistema único, e adaptável ao sistema de saúde pública do Brasil.

6.2 Trabalhos futuros

Como trabalhos futuros existe a possibilidade da ampliação da identificação tanto de microrganismos, incorporando novas regiões e acrescentados outros bancos de dados específicos, para um melhor mapeamento de cepas e espécies tipos.

Há possibilidade de ser desenvolver outros módulos de doenças genéticas específicas cuja identificação possa ser realizada através de sequenciamento e posterior alinhamento contra bancos de dados de referência, proporcionando um perfil exato do diagnóstico de patogenias hereditárias ou não, baseadas em sequenciamento de genes.

Por fim, há possibilidade da implantação desse sistema para a análise da dispersão de patógenos e ou doenças por micro e macro regiões, estendendo para um perfil de dispersão por municípios ou estados.

Referências

- ALLERS, T.; MEVARECH, M. Archaeal genetics - the third way. **Nat Rev Genet**, 2005. v. 6, n. 1, p. 58–73, 2005.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. **J Mol Biol**, 1990. v. 215, n. 3, p. 403–410, 1990.
- ALTSCHUL, S. F. et al. Gapped blast and psi-blast: a new generation of protein database search programs. **Nucleic Acids Res**, 1997. v. 25, n. 17, p. 3389–3402, 1997.
- ANDREI, A.; ZERVOS, M. J. The application of molecular techniques to the study of hospital infection. **Arch Pathol Lab Med**, 2006. v. 130, n. 5, p. 662–668, 2006.
- ANGUS, D. C. et al. Epidemiology of severe sepsis in the united states: analysis of incidence, outcome, and associated costs of care. **Crit Care Med**, 2001. v. 29, n. 7, p. 1303–1310, 2001.
- ARORA, P. K. et al. Chryseomicrobium imtechense gen. nov., sp. nov., a new member of the family planococcaceae. **Int J Syst Evol Microbiol**, 2011. v. 61, p. 1859–1864, 2011.
- ASHELFORD, K. E. et al. At least 1 in 20 16s rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. **Appl Environ Microbiol**, 2005. v. 71, n. 12, p. 7724–7736, 2005.
- AUSUBEL, F. et al. **Current Protocols in Molecular Biology**. [S.l.]: John Wiley & Sons, 2003. ISSN 1934-3647. ISBN 9780471142720.
- BAENA, S. et al. Desulfosoma caldarium gen. nov., sp. nov., a thermophilic sulfate-reducing bacterium from a terrestrial hot spring. **Int J Syst Evol Microbiol**, 2011. v. 61, p. 732–736, 2011.
- BANSAL, A. K. Bioinformatics in microbial biotechnology: a mini review. **Microbial Cell Factories**, 2005. v. 4, n. 19, p. 1–11, 2005.
- BAXEVANIS, A. D.; OUELLETTE, B. F. F. **Bioinformatics: a practical guide to the analysis of genes and proteins**. [S.l.]: John Wiley & Sons, Inc, 2001.
- BEHRENDT, U. et al. Pseudomonas cedrina subsp. fulgida subsp. nov., a fluorescent bacterium isolated from the phyllosphere of grasses; emended description of pseudomonas cedrina and description of pseudomonas cedrina subsp. cedrina subsp. nov. **Int J Syst Evol Microbiol**, 2009. v. 59, n. Pt 6, p. 1331–1335, 2009.
- BONE, R. C. et al. Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. the accp/sccm consensus conference

committee. american college of chest physicians/society of critical care medicine. 1992. **Chest**, 2009. v. 136, n. 5 Suppl, p. e28, 2009.

BOSSHARD, P. P. et al. Comparison of conventional and molecular methods for identification of aerobic catalase-negative gram-positive cocci in the clinical laboratory. **J Clin Microbiol**, 2004. v. 42, n. 5, p. 2065–2073, 2004.

BOSSHARD, P. P. et al. Ribosomal dna sequencing for identification of aerobic gram-positive rods in the clinical laboratory (an 18-month evaluation). **J Clin Microbiol**, 2003. v. 41, n. 9, p. 4134–4140, 2003.

BOYER, S.; FLECHTNER, V.; JOHANSEN, J. Is the 16s-23s rna internal transcribed spacer region a good tool for use in molecular systematics and population genetics? a case study in cyanobacteria. **Molecular Biology and Evolution**, 2001. v. 18, n. 6, p. 1057–1069, 2001.

BRAZAS, M. D. et al. The 2011 bioinformatics links directory update: more resources, tools and databases and features to empower the bioinformatics community. **Nucleic Acids Research**, 2011. v. 39, n. suppl 2, p. W3–W7, 2011.

BROSSETTE, S. E. et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? **Am J Clin Pathol**, 2006. v. 125, n. 1, p. 34–39, 2006.

BRUNAK, S. et al. Nucleotide sequence database policies. **Science**, 2002. v. 298, n. 5597, p. 1333, 2002.

BUTTE, A. J. Translational bioinformatics: coming of age. **J Am Med Inform Assoc**, 2008. v. 15, n. 6, p. 709–714, 2008.

CAMELLO, T. C. F.; AL. et. Nondiphtherial corynebacterium species isolated from clinical specimens of patients in a university hospital. **Braz. j. microbiol**, 2003. v. 34, n. 1, p. 39–44, 2003.

CASE, R. J. et al. Use of 16s rna and rpob genes as molecular markers for microbial ecology studies. **Appl Environ Microbiol**, 2007. v. 73, n. 1, p. 278–288, 2007.

CHAKRAVORTY, S. et al. A detailed analysis of 16s ribosomal rna gene segments for the diagnosis of pathogenic bacteria. **Journal of Microbiological Methods**, 2007. v. 69, n. 2, p. 330–339, 2007. ISSN 0167-7012.

CHRISTEN, R. Global sequencing: A review of current molecular data and new methods available to assess microbial diversity. **Microbes and Enviroments**, 2008. JAPANESE SOC MICROBIAL ECOLOGY, DEPT BIORESOURCE SCIENCE, C/O DR. HIROYUKI OHTA, SEC, IBARAKI UNIV COLLEGE OF AGRICULT, AMI-MACHI, IBARAKI, JAPAN, v. 23, n. 4, p. 253–268, 2008.

CLARRIDGE, J. E. Impact of 16s rna gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. **Clin Microbiol Rev**, 2004. v. 17, n. 4, p. 840–62, table of contents, 2004.

CLOUD, J. L. et al. Evaluation of partial 16s ribosomal dna sequence for identification of norcadia species by using the microseq 500 system with an expanded database. **Journal of Clinical Microbiology**, 2004. v. 42, n. 2, p. 578–584, 2004.

- COLE, J. R. et al. The ribosomal database project (rdp-ii): sequences and tools for high-throughput rRNA analysis. **Nucl. Acids Res.**, 2005. v. 33, n. suppl 1, p. D294–D296, 2005.
- COLE, J. R. et al. The ribosomal database project (rdp-ii): introducing myrdb space and quality controlled public data. **Nucl. Acids Res.**, 2007. v. 35, n. suppl 1, p. D169–D172, 2007.
- COOPER, J. E.; FEIL, E. J. Multilocus sequence typing—what is resolved? **Trends Microbiol.**, 2004. v. 12, n. 8, p. 373–377, 2004.
- DABBOUSSI, F. et al. Taxonomic study of bacteria isolated from lebanese spring waters: proposal for *Pseudomonas cedrella* sp. nov. and *P. orientalis* sp. nov. **Res Microbiol.**, 1999. v. 150, n. 5, p. 303–316, 1999.
- DAUGA, C.; DORÉ, J.; SGHIR, A. Expanding the known diversity and environmental distribution of cultured and uncultured bacteria. **Med Sci (Paris)**, 2005. v. 21, n. 3, p. 290–296, 2005.
- DEGENER, J. E. et al. Nosocomial infection by *Staphylococcus haemolyticus* and typing methods for epidemiological study. **J Clin Microbiol.**, 1994. v. 32, n. 9, p. 2260–2265, 1994.
- DERVISOGLU, E. et al. Peritonitis due to *Pseudomonas putida* in a patient receiving automated peritoneal dialysis. **Infection**, 2008. v. 36, n. 4, p. 379–380, 2008.
- DESANTIS, T. Z. et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. **Appl Environ Microbiol.**, 2006. v. 72, n. 7, p. 5069–5072, 2006.
- DEUTSCHER, D. et al. Implementing an integrated electronic outcomes and electronic health record process to create a foundation for clinical practice improvement. **Phys Ther.**, 2008. v. 88, n. 2, p. 270–285, 2008.
- DEVULDER, G. et al. Bibi, a bioinformatics bacterial identification tool. **J Clin Microbiol.**, 2003. v. 41, n. 4, p. 1785–1787, 2003.
- DO, C. B. et al. Probcons: Probabilistic consistency-based multiple sequence alignment. **Genome Res.**, 2005. v. 15, n. 2, p. 330–340, 2005.
- DUC, L. H. et al. Characterization of *Bacillus* probiotics available for human use. **Appl Environ Microbiol.**, 2004. v. 70, n. 4, p. 2161–2171, 2004.
- EDGAR, R. C. Muscle: a multiple sequence alignment method with reduced time and space complexity. **BMC Bioinformatics**, 2004. v. 5, p. 113, 2004.
- EDGAR, R. C. Muscle: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Res.**, 2004. v. 32, n. 5, p. 1792–1797, 2004.
- ESTEBAN, J. et al. Microbiological characterization and clinical significance of *Corynebacterium amycolatum* strains. **Eur J Clin Microbiol Infect Dis.**, 1999. v. 18, n. 7, p. 518–521, 1999.
- FEIL, E. J. et al. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. **J Bacteriol.**, 2004. v. 186, n. 5, p. 1518–1530, 2004.

FENG, D. F.; DOOLITTLE, R. F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. **J Mol Evol**, 1987. v. 25, n. 4, p. 351–360, 1987.

FERNANDO, B. et al. Benefits and risks of structuring and/or coding the presenting patient history in the electronic health record: systematic review. **BMJ Qual Saf**, 2012. v. 21, n. 4, p. 337–346, 2012.

FIELDS, M. W. et al. Phylogenetic and functional biomarkers as indicators of bacterial community responses to mixed-waste contamination. **Environ Sci Technol**, 2006. v. 40, n. 8, p. 2601–2607, 2006.

FLEISCHMANN, R. D. e. a. Whole-genome random sequencing and assembly of haemophilus influenzae rd. **Science**, 1995. v. 269, n. 5223, p. 496–512, 1995. ISSN 0036-8075.

FUNKE, G. et al. Clinical microbiology of coryneform bacteria. **Clin Microbiol Rev**, 1997. v. 10, n. 1, p. 125–159, 1997.

GALPERIN, M. Y.; FERNANDEZ-SUAREZ, X. M. The 2012 nucleic acids research database issue and the online molecular biology database collection. **Nucleic Acids Research**, 2012. v. 40, n. D1, p. D1–D8, 2012.

GARG, A. X. et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. **JAMA**, 2005. v. 293, n. 10, p. 1223–1238, 2005.

GATSELIS, N. et al. Direct detection of cardiobacterium hominis in serum from a patient with infective endocarditis by broad-range bacterial pcr. **J Clin Microbiol**, 2006. v. 44, n. 2, p. 669–672, 2006.

GIUNTA, A. P. N.; LACERDA, R. A. Inspeção dos programas de controle de infecção hospitalar dos serviços de saúde pela vigilância sanitária: diagnóstico de situação. **Rev Esc Enferm USP**, 2006. v. 40, n. 1, p. 64–70, 2006.

GLASS, J. et al. Essential genes of a minimal bacterium. **Proceedings of the National Academy of Sciences of the United States of America**, 2006. v. 103, n. 2, p. 425–430, 2006. ISSN 0027-8424.

GOTOH, O. Multiple sequence alignment: algorithms and applications. **Adv Biophys**, 1999. v. 36, p. 159–206, 1999.

GRIFFEN, A. L. et al. Core: a phylogenetically-curated 16s rdna database of the core oral microbiome. **PLoS One**, 2011. v. 6, n. 4, p. e19051, 2011.

HARMSSEN, D. et al. Ridom: comprehensive and public sequence database for identification of mycobacterium species. **BMC Infect Dis**, 2003. v. 3, p. 26, 2003.

HARMSSEN, D. et al. Ridom: Ribosomal differentiation of medical micro-organisms database. **Nucleic Acids Res**, 2002. v. 30, n. 1, p. 416–417, 2002.

HARTE, N. et al. Public web-based services from the european bioinformatics institute. **Nucleic Acids Research**, 2004. v. 32, n. suppl 2, p. W3–W9, 2004.

HAUSCHILD, T.; SCHWARZ, S. Differentiation of staphylococcus sciuri strains isolated from free-living rodents and insectivores. **J Vet Med B Infect Dis Vet Public Health**, 2003. v. 50, n. 5, p. 241–246, 2003.

- HAYNES, R. B.; WILCZYNSKI, N. L.; , C. C. D. S. S. C. C. D. S. S. S. R. T. . Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: methods of a decision-maker-researcher partnership systematic review. **Implement Sci**, 2010. v. 5, p. 12, 2010.
- HERTZ, G. Z.; STORMO, G. D. Identifying dna and protein patterns with statistically significant alignments of multiple sequences. **Bioinformatics**, 1999. v. 15, n. 7-8, p. 563–577, 1999.
- HIROSAWA, M. et al. Comprehensive study on iterative algorithms of multiple sequence alignment. **Comput Appl Biosci**, 1995. v. 11, n. 1, p. 13–18, 1995.
- HOTA, B. et al. Formulation of a model for automating infection surveillance: algorithmic detection of central-line associated bloodstream infection. **J Am Med Inform Assoc**, 2010. v. 17, n. 1, p. 42–48, 2010.
- IMHOFF, J. F.; CAUMETTE, P. Recommended standards for the description of new species of anoxygenic phototrophic bacteria. **Int J Syst Evol Microbiol**, 2004. v. 54, n. Pt 4, p. 1415–1421, 2004.
- JANDA, J. M.; ABBOTT, S. L. 16s rna gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. **J Clin Microbiol**, 2007. v. 45, n. 9, p. 2761–2764, 2007.
- JENSEN L. B., M.; WEBSTER, J.; STRAUS, N. Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal dna spacer polymorphisms. **Applied and Environmental Microbiology**, 1993. v. 59, n. 4, p. 945–952, 1993.
- KANN, M. G. Advances in translational bioinformatics: computational approaches for the hunting of disease genes. **Brief Bioinform**, 2010. v. 11, n. 1, p. 96–110, 2010.
- KANZ, C. et al. The embl nucleotide sequence database. **Nucleic Acids Res**, 2005. v. 33, n. Database issue, p. D29–D33, 2005.
- KATOH, K. et al. Mafft version 5: improvement in accuracy of multiple sequence alignment. **Nucleic Acids Res**, 2005. v. 33, n. 2, p. 511–518, 2005.
- KATOH, K.; TOH, H. Recent developments in the mafft multiple sequence alignment program. **Brief Bioinform**, 2008. v. 9, n. 4, p. 286–298, 2008.
- KAZUMI, Y.; MAEDA, S.; SUGAWARA, I. Identification of mycobacteria by sequencing of rpob gene and 16s rna. **Kekkaku**, 2006. v. 81, n. 9, p. 551–558, 2006.
- KILBREDGE, P. M.; CLASSEN, D. C. The informatics opportunities at the intersection of patient safety and clinical informatics. **J Am Med Infor Assoc**, 2008. v. 15, n. 4, p. 397–407, 2008.
- KIMURA, Y. et al. Acute peritonitis due to corynebacterium ulcerans in a patient receiving continuous ambulatory peritoneal dialysis: a case report and literature review. **Clin Exp Nephrol**, 2011. v. 15, n. 1, p. 171–174, 2011.
- KLOOS, W. E.; SCHLEIFER, K. H.; SMITH, R. F. Characterization of staphylococcus sciuri sp. nov. and its subspecies. **Int J Syst Bacteriol**, 1976. v. 26, n. 1, p. 22–37, 1976.

Kämpfer, P.; Glaeser, S. P. Prokaryotic taxonomy in the sequencing era—the polyphasic approach revisited. **Environ Microbiol**, 2012. v. 14, n. 2, p. 291–317, 2012.

Kodama, Y. et al. The dna data bank of japan launches a new resource, the ddbj omics archive of functional genomics experiments. **Nucleic Acids Res**, 2012. v. 40, n. Database issue, p. D38–D42, 2012.

Lacerda, R. A.; Jouclas, V. M. G.; Egrý, E. Y. Infecções hospitalares no brasil: Ações governamentais para o seu controle enquanto expressão de políticas sociais na área de saúde. **Rev. Esc. Enferm**, 1996. v. 30, n. 1, 1996.

Larkin, M. A. et al. Clustal w and clustal x version 2.0. **Bioinformatics**, 2007. v. 23, n. 21, p. 2947–2948, 2007.

Liu, Z. et al. Accurate taxonomy assignments from 16s rRNA sequences produced by highly parallel pyrosequencers. **Nucleic Acids Res**, 2008. v. 36, n. 18, p. e120, 2008.

Logan, N. A. et al. Polyphasic identification of bacillus and brevibacillus strains from clinical, dairy and industrial specimens and proposal of brevibacillus invocatus sp. nov.. **Int J Syst Evol Microbiol**, 2002. v. 52, n. Pt 3, p. 953–966, May 2002.

Lopez, P.; Baptiste, E. Molecular phylogeny: reconstructing the forest. **C R Biol**, 2009. v. 332, n. 2-3, p. 171–182, 2009.

Loureiro, M. M. et al. Study of multi-drug resistant microorganisms isolated from blood cultures of hospitalized newborns in rio de janeiro city, brazil. **Brazilian Journal of Microbiology**, 2002. v. 33, n. 1, p. 73–78, 2002.

Loy, A. et al. probabase—an online resource for rRNA-targeted oligonucleotide probes: new features 2007. **Nucleic Acids Res**, 2007. v. 35, n. Database issue, p. D800–D804, 2007.

Lu, X.-X. et al. 16s rRNA gene sequencing for pathogen identification from clinical specimens. **Zhonghua Yi Xue Za Zhi**, 2008. v. 88, n. 2, p. 123–126, 2008.

MacFarlane, L.; Oppenheim, B. A.; Lorrigan, P. Septicaemia and septic arthritis due to pseudomonas putida in a neutropenic patient. **J Infect**, 1991. v. 23, n. 3, p. 346–347, 1991.

Mahenthiralingam, E.; Baldwin, A.; Dowson, C. G. Burkholderia cepacia complex bacteria: opportunistic pathogens with important natural biology. **J Appl Microbiol**, 2008. v. 104, n. 6, p. 1539–1551, 2008.

Maidak, B. L. et al. The rdp-ii (ribosomal database project). **Nucl. Acids Res.**, 2001. v. 29, n. 1, p. 173–174, 2001.

Maidak, B. L. et al. A new version of the rdp (ribosomal database project). **Nucleic Acids Res**, 1999. v. 27, n. 1, p. 171–173, 1999.

Marseglia, G. L. et al. Efficacy of bacillus clausii spores in the prevention of recurrent respiratory infections in children: a pilot study. **Ther Clin Risk Manag**, 2007. v. 3, n. 1, p. 13–17, 2007.

Martínez-Murcia, A. J.; Antón, A. I.; Rodríguez-Valera, F. Patterns of sequence variation in two regions of the 16s rRNA multigene family of escherichia coli. **Int J Syst Bacteriol**, 1999. v. 49 Pt 2, p. 601–610, 1999.

- MCKIBBEN, L. et al. Ensuring rational public reporting systems for health care-associated infections: systematic literature review and evaluation recommendations. **Am J Infect Control**, 2006. v. 34, n. 3, p. 142–149, 2006.
- MIGNARD, S.; FLANDROIS, J. P. 16s rRNA sequencing in routine bacterial identification: a 30-J **Microbiol Methods**, 2006. v. 67, n. 3, p. 574–581, 2006.
- MONTO, A. S.; WHITLEY, R. J. Seasonal and pandemic influenza: a 2007 update on challenges and solutions. **Clin Infect Dis**, 2008. v. 46, n. 7, p. 1024–1031, 2008.
- MORGENSTERN, B.; DRESS, A.; WERNER, T. Multiple DNA and protein sequence alignment based on segment-to-segment comparison. **Proc Natl Acad Sci U S A**, 1996. v. 93, n. 22, p. 12098–12103, 1996.
- MORIYA, T. et al. *Calditerricola satsumensis* gen. nov., sp. nov. and *calditerricola yamamurae* sp. nov., extreme thermophiles isolated from a high-temperature compost. **Int J Syst Evol Microbiol**, 2011. v. 61, p. 631–639, 2011.
- MURATA, M. Three-way Needleman–Wunsch algorithm. **Methods Enzymol**, 1990. v. 183, p. 365–375, 1990.
- NEMERGUT, D. R. et al. Global patterns in the biogeography of bacterial taxa. **Environ Microbiol**, 2011. v. 13, n. 1, p. 135–144, 2011.
- NISHIMURA, Y. et al. Radiation sensitivities of *Acinetobacter* strains isolated from clinical sources. **J Basic Microbiol**, 1994. v. 34, n. 5, p. 357–360, 1994.
- NOTREDAME, C. Recent progress in multiple sequence alignment: a survey. **Pharmacogenomics**, 2002. v. 3, n. 1, p. 131–144, 2002.
- OLSEN, G. J. et al. The ribosomal database project. **Nucleic Acids Res**, 1992. v. 20 Suppl, p. 2199–2200, 1992.
- OLSEN, G. J.; WOESE, C. R. Ribosomal rRNA: a key to phylogeny. **FASEB J**, 1993. v. 7, n. 1, p. 113–123, 1993.
- PACE, N. R. A molecular view of microbial diversity and the biosphere. **Science**, 1997. v. 276, n. 5313, p. 734–740, 1997.
- PAGANI, I. et al. The genomes online database (gold) v.4: status of genomic and metagenomic projects and their associated metadata. **Nucleic Acids Research**, 2012. v. 40, n. D1, p. D571–D579, 2012.
- PATEL, J. B. 16s rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. **Mol Diagn**, 2001. v. 6, n. 4, p. 313–321, 2001.
- PATEL, J. B. et al. Sequence-based identification of *Mycobacterium* species using the microseq 500 16s rDNA bacterial identification system. **Journal of Clinical Microbiology**, 2000. v. 38, n. 1, p. 246–251, 2000.
- PATEL, R. et al. Determination of 16s rRNA sequences of enterococci and application to species identification of nonmotile enterococcus gallinarum isolates. **J Clin Microbiol**, 1998. v. 36, n. 11, p. 3399–3407, 1998.
- PEI, J.; GRISHIN, N. V. Promals: towards accurate multiple sequence alignments of distantly related proteins. **Bioinformatics**, 2007. v. 23, n. 7, p. 802–808, 2007.

PETTI, C. A. Detection and identification of microorganisms by gene amplification and sequencing. **Medical Microbiology**, 2007. v. 44, p. 1108–1114, 2007.

PITTET, D. Infection control and quality health care in the new millenium. **American Journal of Infection Control**, 2005. v. 33, n. 5, p. 258–267, 2005.

PRONOVOST, P. et al. How will we know patients are safer? an organization-wide approach to measuring and improving safety. **Crit Care Med**, 2006. v. 34, n. 7, p. 1988–1995, 2006.

PRONOVOST, P. J. et al. Intensive care unit physician staffing: financial modeling of the leapfrog standard. **Crit Care Med**, 2006. v. 34, n. 3 Suppl, p. S18–S24, 2006.

PRONOVOST, P. J. et al. Toward learning from patient safety reporting systems. **J Crit Care**, 2006. v. 21, n. 4, p. 305–315, 2006.

PRUESSE, E. et al. Silva: a comprehensive online resource for quality checked and aligned ribosomal rna sequence data compatible with arb. **Nucleic Acids Res**, 2007. v. 35, n. 21, p. 7188–7196, 2007.

RANDELL, R. et al. Effects of computerized decision support systems on nursing performance and patient outcomes: a systematic review. **J Health Serv Res Policy**, 2007. v. 12, n. 4, p. 242–249, 2007.

RASTOGI, R. et al. Visualization of ribosomal rna operon copy number distribution. **BMC Microbiol**, 2009. v. 9, p. 208, 2009.

ROLSTON, K. V. I. et al. Nonfermentative gram-negative bacilli in cancer patients: increasing frequency of infection and antimicrobial susceptibility of clinical isolates to fluoroquinolones. **Diagn Microbiol Infect Dis**, 2005. v. 51, n. 3, p. 215–218, Mar 2005.

ROTHBERG, M. B.; BENJAMIN, E. M.; LINDENAUER, P. K. Public reporting of hospital quality: recommendations to benefit patients and hospitals. **J Hosp Med**, 2009. v. 4, n. 9, p. 541–545, 2009.

SAHA, R.; JAIN, S.; KAUR, I. R. Metallo beta-lactamase producing pseudomonas species—a major cause of concern among hospital associated urinary tract infection. **J Indian Med Assoc**, 2010. v. 108, n. 6, p. 344–348, 2010.

SAHA, S. C.; CHOPADE, B. A. Radiation sensitivity of acinetobacter spp. and their redicidation for preservation of meat at low temperature. **Bangladesh Med Res Counc Bull**, 2009. v. 35, n. 2, p. 33–40, 2009.

SANKOFF, D. Matching sequences under deletion-insertion constraints. **Proc Natl Acad Sci U S A**, 1972. v. 69, n. 1, p. 4–6, 1972.

SARKAR, I. N. et al. Translational bioinformatics: linking knowledge across biological and clinical realms. **Journal of the American Medical Informatics Association**, 2011. BMJ Publishing Group Ltd, v. 18, n. 4, p. 354–357, 2011.

SCHLEIFER, K. H. Classification of bacteria and archaea: past, present and future. **Syst Appl Microbiol**, 2009. v. 32, n. 8, p. 533–542, 2009.

SEHULSTER, L. M. et al. **Centers for Disease Control and Prevention Healthcare Infection Control Practices Advisory Committee (HICPAC): Guidelines for Environmental Infection Control in Health-Care Facilities.** [S.l.], 2003.

SEIFERT, H. et al. Distribution of acinetobacter species on human skin: comparison of phenotypic and genotypic identification methods. **J Clin Microbiol**, 1997. v. 35, n. 11, p. 2819–2825, 1997.

SHENG, W. H. et al. Comparative impact of hospital-acquired infections on medical costs, length of hospital stay and outcome between community hospitals and medical centres. **J Hosp Infect**, 2005. v. 59, n. 3, p. 205–214, 2005.

SHIDA, O. et al. Proposal for two new genera, *brevibacillus* gen. nov. and *aneurinibacillus* gen. nov. **Int J Syst Bacteriol**, 1996. v. 46, n. 4, p. 939–946, 1996.

SIDHU, A. S.; CHEN, J. **Biological Database Modeling.** [S.l.]: Artech House, 2008.

SINGH, A. et al. Application of molecular techniques to the study of hospital infection. **Clinical Microbiology Reviews**, July 2006. v. 19, n. 3, p. 512–530, July 2006.

SINGHI, S. C.; BARANWAL, A. Probiotic use in the critically ill. **Indian J Pediatr**, 2008. v. 75, n. 6, p. 621–627, 2008.

SMITH, T. F.; WATERMAN, M. S. Identification of common molecular subsequences. **J Mol Biol**, 1981. v. 147, n. 1, p. 195–197, 1981.

SPANG, R.; VINGRON, M. Limits of homology detection by pairwise sequence comparison. **Bioinformatics**, 2001. v. 17, n. 4, p. 338–342, 2001.

STEPANOVIC, S. et al. Isolation of members of the *staphylococcus sciuri* group from urine and their relationship to urinary tract infections. **J Clin Microbiol**, 2003. v. 41, n. 11, p. 5262–5264, 2003.

STRUELENS, M. J.; DENIS, O.; RODRIGUEZ-VILLALOBOS, H. Microbiology of nosocomial infections: progress and challenges. **Microbes Infect**, 2004. v. 6, n. 11, p. 1043–1048, 2004.

TAMURA, K. et al. Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Mol Biol Evol**, 2011. v. 28, n. 10, p. 2731–2739, 2011.

TANG, Y.-W.; PROCOP, G. W.; PERSING, D. H. Molecular diagnostics of infectious diseases. **Clinical Chemistry**, 1997. v. 43, n. 11, p. 2021–2038, 1997.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res**, 1994. v. 22, n. 22, p. 4673–4680, 1994.

THOMPSON, J. D.; PLEWNIAK, F.; POCH, O. Balibase: a benchmark alignment database for the evaluation of multiple alignment programs. **Bioinformatics**, 1999. v. 15, n. 1, p. 87–88, 1999.

- THOMPSON, J. D.; PLEWNIAK, F.; POCH, O. A comprehensive comparison of multiple sequence alignment programs. **Nucleic Acids Res**, 1999. v. 27, n. 13, p. 2682–2690, 1999.
- TREVIÑO, M. et al. Nosocomial infection by vim-2 metallo-beta-lactamase-producing pseudomonas putida. **J Med Microbiol**, 2010. v. 59, n. Pt 7, p. 853–855, 2010.
- TURENNE, C. Y. et al. Necessity of quality-controlled 16s rna gene sequence databases: identifying nontuberculous mycobacterium species. **J Clin Microbiol**, 2001. v. 39, n. 10, p. 3637–3648, 2001.
- URWIN, R.; MAIDEN, M. C. J. Multi-locus sequence typing: a tool for global epidemiology. **Trends Microbiol**, 2003. v. 11, n. 10, p. 479–487, 2003.
- WALLACE, I. M.; BLACKSHIELDS, G.; HIGGINS, D. G. Multiple sequence alignments. **Curr Opin Struct Biol**, 2005. v. 15, n. 3, p. 261–266, 2005.
- WALLET, F. et al. Peritonitis due to staphylococcus sciuri in a patient on continuous ambulatory peritoneal dialysis. **Scand J Infect Dis**, 2000. v. 32, n. 6, p. 697–698, 2000.
- WANG, Q. et al. Naive bayesian classifier for rapid assignment of rna sequences into the new bacterial taxonomy. **Appl Environ Microbiol**, 2007. v. 73, n. 16, p. 5261–5267, 2007.
- WATERHOUSE, A. M. et al. Jalview version 2—a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, 2009. v. 25, n. 9, p. 1189–1191, 2009.
- WISPLINGHOFF, H. et al. Nosocomial bloodstream infections caused by acinetobacter species in united states hospitals: clinical features, molecular epidemiology, and antimicrobial susceptibility. **Clin Infect Dis**, 2000. v. 31, n. 3, p. 690–697, 2000.
- WOESE, C. R. et al. Conservation of primary structure in 16s ribosomal rna. **Nature**, 1975. v. 254, n. 5495, p. 83–86, 1975.
- WONG, J. S. J. et al. Corynebacterium accolens-associated pelvic osteomyelitis. **J Clin Microbiol**, 2010. v. 48, n. 2, p. 654–655, 2010.
- WOO, P. C. Y. et al. Then and now: use of 16s rDNA gene sequencing for bacterial identification and discovery of novel bacteria in clinical microbiology laboratories. **Clin Microbiol Infect**, 2008. v. 14, n. 10, p. 908–934, 2008.
- WOO, P. C. Y. et al. Guidelines for interpretation of 16s rna gene sequence-based results for identification of medically important aerobic gram-positive bacteria. **J Med Microbiol**, 2009. v. 58, n. Pt 8, p. 1030–1036, 2009.
- WUYTS, J.; PERRIERE, G.; PEER, Y. Van de. The european ribosomal rna database. **Nucl. Acids Res.**, 2004. v. 32, n. suppl 1, p. D101–D103, 2004.
- YEANG, C.-H. et al. Detecting the coevolution of biosequences—an example of rna interaction prediction. **Mol Biol Evol**, 2007. v. 24, n. 9, p. 2119–2131, 2007.
- YOSHINO, Y. et al. Pseudomonas putida bacteremia in adult patients: five case reports and a review of the literature. **J Infect Chemother**, 2011. v. 17, n. 2, p. 278–282, Apr 2011.

ZDOBNOV, E. M. et al. The ebi srs server-recent developments. **Bioinformatics**, 2002. v. 18, n. 2, p. 368–373, 2002.

1 Apêndice - Formulário de coleta de dados do paciente



UNIVERSIDADE FEDERAL DE ALAGOAS – UFAL
CAMPUS ARAPIRACA
LABORATÓRIO DE MICROBIOLOGIA MÉDICA E BIOLOGIA MOLECULAR

FORMULÁRIO DE IDENTIFICAÇÃO DE PACIENTE-AMOSTRA

<u>PACIENTE</u>	Código da Amostra:
Nome: _____ - Sexo: ☐ M ☐ F Registro (Prontuário): _____ - Data de Nascimento: ____/____/____ - CPF: _____ Data da Coleta: ____/____/____ - Hora: _____ - Data da Entrada: ____/____/____ Setor de Coleta: ☐ U.T.I. _____ ☐ Enfermaria _____ ☐ Centro Cirúrgico _____ ☐ Ambulatório _____ ☐ Outro _____	
Dados Clínicos: _____ Dados Epidemiológicos: _____ ☐ Acidente automobilístico ☐ Contato com animais ☐ Doença ocupacional ☐ Outro _____ Dados Laboratoriais: _____ Provável origem do processo infeccioso: ☐ Comunitária ☐ Hospitalar Em Procedimento Invasivo: _____ ☐ Sonda Vesical ☐ Cateter ☐ Traqueostomia ☐ Alimentação parenteral ☐ Cirurgia ☐ Outro _____ Existe infecção em outra topografia? ☐ Não ☐ Sim _____ Fez uso de antibióticos nos últimos 10 dias? ☐ Não ☐ Sim _____ Comprometimento Imunológico? ☐ Não ☐ Sim _____ É paciente transferido ou de alta de outro hospital nos últimos 30 dias? ☐ Não ☐ Sim _____ É portador, colonizado ou infectado de bactérias multirresistentes? ☐ Não ☐ Sim _____ Para Urocultura informar: ☐ Sintomático ☐ Assintomático Nome e identificação do coletador: _____ Observações: _____	
<u>AMOSTRA</u> Hemocultura: ☐ Sangue Periférico ☐ Sangue arterial Urocultura: ☐ Jato médio ☐ Sonda vesical de demora ☐ Sonda de alívio ☐ Punção Suprapúbica Trato Respiratório: ☐ Orofaringe ☐ Escarro ☐ Aspirado traqueal ☐ Lavado bronco-alveolar ☐ Escovado brônquico ☐ Outro _____ Lesões, secreções, abscessos: ☐ - Descrever topografia _____ Via de obtenção do material: ☐ Swab ☐ Punção ☐ Drenagem ☐ Outro _____ Líquidos cavitários: ☐ Líquor ☐ Líquido pleural ☐ Líquido Ascítico ☐ Líq. Sinovial ☐ Líq. Pericárdico Material: ☐ Ponta de cateter ☐ Próteses ☐ Outro _____ Observações: _____	

Microrganismo Isolado:

Microbiológico Tradicional _____

Diagnóstico Molecular _____

Observações:

Código Bacterioteca:

--

Antibiograma

Antibióticos Beta-Lactâmicos	Perfil		Antibióticos Não Beta-Lactâmicos	Perfil

* R – Resistente; S – Sensível; I – Intermediário; NT – Não Testado.

Observações:

2 Apêndice - Termo de compromisso livre e esclarecido

Apêndice – Termo de Consentimento Livre e Esclarecido.

Termo de Consentimento Livre e Esclarecido (T.C.L.E.)

“O respeito devido à dignidade humana exige que toda pesquisa se processe após consentimento livre e esclarecido dos sujeitos, indivíduos ou grupos que por si e/ou por seus representantes legais manifestem a sua anuência à participação na pesquisa.” (Resolução. nº 196/96-IV, do Conselho Nacional de Saúde)

Eu, ..., tendo sido convidado(a) a participar como voluntário(a) do estudo “Vigilância Epidemiológica na Unidade de Emergência do Agreste: Implantação de Serviço de Epidemiologia Molecular e Criação de Banco de Dados Informatizado para sequências de rDNA 16S de bactérias patogênicas.”, recebi do Sr. Prof. Dr. Tiago Gomes de Andrade, do Campus Arapiraca da Universidade Federal de Alagoas-UFAL, responsáveis por sua execução, as seguintes informações que me fizeram entender sem dificuldades e sem dúvidas os seguintes aspectos:

1. Que o estudo se destina a investigar quais os agentes infecciosos (bactérias) responsáveis pelos casos de infecção hospitalar na Unidade de Emergência do Agreste (UEA). A infecção hospitalar é caracterizada pela infecção adquirida durante o período de hospitalização.
2. Que a importância deste estudo é a de melhorar os serviços de saúde e a qualidade de atendimento dos pacientes.
3. Que os resultados que se desejam alcançar são os seguintes: fazer um levantamento dos tipos de bactérias que causam doença (infecção hospitalar) na UEA; criar um serviço especializado e moderno para identificação e controle destas bactérias. Os resultados desta pesquisa podem ajudar no controle das infecções hospitalares, criando condições para que outras pessoas tenham menores riscos de contaminação.
4. Que esse estudo começará em março de 2008 e terminará em março/2010.
5. Que o estudo será feito da seguinte maneira: Serão coletadas amostras de material biológico, como sangue, urina ou secreções, dos pacientes internos na UEA. A coleta de sangue será feita em um dos braços por punção venosa (retirada de sangue pela veia do braço com agulha) realizada com material descartável por profissional treinado. Serão retirados, se necessário, no máximo, 15 mL do meu sangue. Este volume de sangue é cerca de 30 (trinta) vezes menor do que o volume de sangue normalmente doado quando um indivíduo doa sangue para bancos de sangue. Se necessário, a coleta de secreções (“pus”, “catarro”) será feita com materiais plásticos adequados, de maneira indolor (sem causar dor). Se necessário, a coleta de urina será feita em frascos de coleta. Estas amostras serão submetidas à análise em laboratório para identificação da bactéria causadora da infecção, utilizando métodos modernos de biologia molecular e análise de informática.
6. Que eu participarei das seguintes etapas: Participarei durante a coleta através da doação de material biológico (sangue, urina ou secreções). Com isto, encerra-se a minha participação neste estudo.
7. Que os outros meios conhecidos para se obter os mesmos resultados são os seguintes: não existem outros meios para se obter os mesmos resultados.
8. Que os incômodos que poderei sentir com a minha participação são os seguintes: O desconforto da coleta de sangue é dor leve e passageira e/ ou possível aparecimento de mancha roxa no local.
9. Que os possíveis riscos à minha saúde física e mental são: A participação neste estudo não trará nenhum risco à minha saúde física e mental.
10. Que deverei contar com a seguinte assistência: Garantia de que serei devidamente acompanhado (a) e assistido (a) durante todo o período de participação no projeto, sem gastos adicionais, bem como de que será garantida a continuidade do tratamento, após a conclusão dos trabalhos da

pesquisa, sendo responsável por ela: a equipe multidisciplinar de saúde da Unidade de Emergência do Agreste.

11. Que os benefícios que deverei esperar com a minha participação, mesmo que não diretamente são: tratamento adequado para eliminar a possível bactéria responsável pela infecção e o aperfeiçoamento dos serviços de identificação e controle de infecção hospitalar.
12. Que a minha participação será acompanhada do seguinte modo: serei acompanhado pela equipe de saúde e pelo pesquisador responsável durante a coleta de material e pela equipe médica da UEA durante o tratamento.
13. Que, sempre que desejar, serão fornecidos esclarecimentos sobre cada uma das etapas do estudo.
14. Que, a qualquer momento, eu poderei recusar a continuar participando do estudo e, também, que eu poderei retirar este meu consentimento, sem que isso me traga qualquer penalidade ou prejuízo.
15. Que as informações conseguidas através da minha participação não permitirão a identificação da minha pessoa, exceto aos responsáveis pelo estudo, e que a divulgação das mencionadas informações só será feita entre os profissionais estudiosos do assunto.

Finalmente, tendo eu compreendido perfeitamente tudo o que me foi informado sobre a minha participação no mencionado estudo e estando consciente dos meus direitos, das minhas responsabilidades, dos riscos e dos benefícios que a minha participação implicam, concordo em dele participar e para isso eu DOU O MEU CONSENTIMENTO SEM QUE PARA ISSO EU TENHA SIDO FORÇADO OU OBRIGADO.

Endereço do(a) participante-voluntário(a)

Domicílio: (rua, praça, conjunto):
Bloco: /Nº: /Complemento:
Bairro: /CEP/Cidade: /Telefone:
Ponto de referência:

Contato de urgência: Sr(a).

Domicílio: (rua, praça, conjunto):
Bloco: /Nº: /Complemento:
Bairro: /CEP/Cidade: /Telefone:
Ponto de referência:

Endereço do(s) responsável(is) pela pesquisa:

Instituição: Universidade Federal de Alagoas-UFAL Campus Arapiraca
Endereço: Rodovia AL 115, Km 6.5
CEP: 57300-970. Arapiraca-AL
Telefones p/contato: 82- 99930097

ATENÇÃO: Para informar ocorrências irregulares ou danosas durante a sua participação no estudo, dirija-se ao:

Comitê de Ética em Pesquisa da Universidade Federal de Alagoas:

Prédio da Reitoria, sala do C.O.C. , Campus A. C. Simões, Cidade Universitária

Telefone: 3214-1053

Arapiraca,

(Assinatura ou impressão datiloscópica do(a) voluntário(a) ou responsável legal - Rubricar as demais folhas)	Nome e Assinatura do responsável pelo estudo (Rubricar as demais páginas)

3 Apêndice - Sequências depositadas no NCBI

Apêndice - Números de Acesso das Sequências Depositadas no GenBank e Identificação dos Microrganismos.

ID NCBI/GenBank	Descrição das Sequências obtidas
JF919832.1	Acinetobacter baumannii strain 11-7-8 16S ribosomal RNA gene, partial sequence
JF919833.1	Acinetobacter baumannii strain 12-7-8 16S ribosomal RNA gene, partial sequence
JF920005.1	Staphylococcus epidermidis strain 13-10-9 16S ribosomal RNA gene, partial sequence
JF919989.1	Staphylococcus aureus strain 17-9-8 16S ribosomal RNA gene, partial sequence
JF919983.1	Pseudomonas putida strain 13-3-1-9 16S ribosomal RNA gene, partial sequence
JF919982.1	Pseudomonas cedrina strain 13-11-1-8 16S ribosomal RNA gene, partial sequence
JF919874.1	Corynebacterium striatum strain 6-2-10 16S ribosomal RNA gene, partial sequence
JF919937.1	Pseudomonas aeruginosa strain 1-4-1-8 16S ribosomal RNA gene, partial sequence
JF919932.1	Proteus mirabilis strain 1-4-2-8 16S ribosomal RNA gene, partial sequence
JF919848.1	Acinetobacter baumannii strain 15-6-1-9 16S ribosomal RNA gene, partial sequence
JF919968.1	Pseudomonas aeruginosa strain 15-6-2-9 16S ribosomal RNA gene, partial sequence
JF919969.1	Pseudomonas aeruginosa strain 16-6-1-9 16S ribosomal RNA gene, partial sequence
JF919849.1	Acinetobacter baumannii strain 16-6-2-9 16S ribosomal RNA gene, partial sequence
JF919890.1	Enterobacter hormaechei strain 9-9-8 16S ribosomal RNA gene, partial sequence
JF919905.1	Klebsiella pneumoniae strain 4-10-8 16S ribosomal RNA gene, partial sequence
JF919991.1	Staphylococcus aureus strain 10-10-8 16S ribosomal RNA gene, partial sequence
JF919992.1	Staphylococcus aureus strain 13-10-8 16S ribosomal RNA gene, partial sequence
JF920025.1	Klebsiella sp. 6-11-8 16S ribosomal RNA gene, partial sequence
JF919954.1	Pseudomonas aeruginosa strain 7-11-1-8 16S ribosomal RNA gene, partial sequence
JF919837.1	Acinetobacter baumannii strain 7-11-2-8 16S ribosomal RNA gene, partial sequence
JF919994.1	Staphylococcus aureus strain 7-11-3-8 16S ribosomal RNA gene, partial sequence
JF919955.1	Pseudomonas aeruginosa strain 8-11-8 16S ribosomal RNA gene, partial sequence
JF919950.1	Pseudomonas aeruginosa strain 29-10-8 16S ribosomal RNA gene, partial sequence
JF920008.1	Staphylococcus haemolyticus strain 1-11-1-8 16S ribosomal RNA gene, partial sequence
JF919953.1	Pseudomonas aeruginosa strain 3-11-8 16S ribosomal RNA gene, partial sequence
JF919870.1	Brevibacillus agri strain 10-11-8 16S ribosomal RNA gene, partial sequence
JF919993.1	Staphylococcus aureus strain 4-11-8 16S ribosomal RNA gene, partial sequence
JF920029.1	Enterobacteriaceae bacterium 12-11-1-8 16S ribosomal RNA gene, partial sequence
JF919982.1	Pseudomonas cedrina strain 13-11-1-8 16S ribosomal RNA gene, partial sequence
JF919908.1	Klebsiella pneumoniae strain 9-3-2-9 16S ribosomal RNA gene, partial sequence
JF919869.1	Bacillus clausii strain 5-3-1-9 16S ribosomal RNA gene, partial sequence
JF919957.1	Pseudomonas aeruginosa strain 5-3-2-9 16S ribosomal RNA gene, partial sequence
JF919997.1	Staphylococcus aureus strain 18-9-1-9 16S ribosomal RNA gene, partial sequence
JF919911.1	Klebsiella pneumoniae strain 19-9-9 16S ribosomal RNA gene, partial sequence
JF919881.1	Escherichia coli strain 20-10-1-9 16S ribosomal RNA gene, partial sequence
JF919919.1	Klebsiella pneumoniae strain 20-10-2-9 16S ribosomal RNA gene, partial sequence
JF919958.1	Pseudomonas aeruginosa strain 8-3-9 16S ribosomal RNA gene, partial sequence
JF919907.1	Klebsiella pneumoniae strain 6-3-9 16S ribosomal RNA gene, partial sequence
JF919839.1	Acinetobacter baumannii strain 3-3-9 16S ribosomal RNA gene, partial sequence
JF919934.1	Proteus mirabilis strain 42-9-9 16S ribosomal RNA gene, partial sequence
JF919878.1	Escherichia coli strain 1-3-9 16S ribosomal RNA gene, partial sequence

JF919873.1	<i>Corynebacterium striatum</i> strain 2-3-1-9 16S ribosomal RNA gene, partial sequence
JF919838.1	<i>Acinetobacter baumannii</i> strain 2-3-2-9 16S ribosomal RNA gene, partial sequence
JF919842.1	<i>Acinetobacter baumannii</i> strain 4-5-9 16S ribosomal RNA gene, partial sequence
JF919983.1	<i>Pseudomonas putida</i> strain 13-3-1-9 16S ribosomal RNA gene, partial sequence
JF919840.1	<i>Acinetobacter baumannii</i> strain 13-3-2-9 16S ribosomal RNA gene, partial sequence
JF919893.1	<i>Enterobacter hormaechei</i> strain 13-3-3-9 16S ribosomal RNA gene, partial sequence
JF919897.1	<i>Klebsiella granulomatis</i> strain 21-3-9 16S ribosomal RNA gene, partial sequence
JF919963.1	<i>Pseudomonas aeruginosa</i> strain 11-5/9 16S ribosomal RNA gene, partial sequence
JF919967.1	<i>Pseudomonas aeruginosa</i> strain 6-6-9 16S ribosomal RNA gene, partial sequence
JF919961.1	<i>Pseudomonas aeruginosa</i> strain 5-5-1-9 16S ribosomal RNA gene, partial sequence
JF919843.1	<i>Acinetobacter baumannii</i> strain 5-5-2-9 16S ribosomal RNA gene, partial sequence
JF919909.1	<i>Klebsiella pneumoniae</i> strain 9-5-9 16S ribosomal RNA gene, partial sequence
JF919841.1	<i>Acinetobacter baumannii</i> strain 1-5-9 16S ribosomal RNA gene, partial sequence
JF919960.1	<i>Pseudomonas aeruginosa</i> strain 3-5-2-9 16S ribosomal RNA gene, partial sequence
JF920006.1	<i>Serratia marcescens</i> strain 5-6-1-9 16S ribosomal RNA gene, partial sequence
JF919966.1	<i>Pseudomonas aeruginosa</i> strain 5-6-2-9 16S ribosomal RNA gene, partial sequence
JF920007.1	<i>Serratia marcescens</i> strain 19-6-1-9 16S ribosomal RNA gene, partial sequence
JF919965.1	<i>Pseudomonas aeruginosa</i> strain 1-6-1-9 16S ribosomal RNA gene, partial sequence
JF919985.1	<i>Providencia stuartii</i> strain 1-6-2-9 16S ribosomal RNA gene, partial sequence
JF919844.1	<i>Acinetobacter baumannii</i> strain 2-6-1-9 16S ribosomal RNA gene, partial sequence
JF919995.1	<i>Staphylococcus aureus</i> strain 2-6-2-9 16S ribosomal RNA gene, partial sequence
JF919846.1	<i>Acinetobacter baumannii</i> strain 9-6-1-9 16S ribosomal RNA gene, partial sequence
JF919996.1	<i>Staphylococcus aureus</i> strain 9-6-2-9 16S ribosomal RNA gene, partial sequence
JF920014.1	<i>Staphylococcus saprophyticus</i> strain 18-6-9 16S ribosomal RNA gene, partial sequence
JF919984.1	<i>Providencia stuartii</i> strain 2-5-2-9 16S ribosomal RNA gene, partial sequence
JF919959.1	<i>Pseudomonas aeruginosa</i> strain 2-5-1-9 16S ribosomal RNA gene, partial sequence
JF919913.1	<i>Klebsiella pneumoniae</i> strain 21-9-9 16S ribosomal RNA gene, partial sequence
JF919850.1	<i>Acinetobacter baumannii</i> strain 5-9-9 16S ribosomal RNA gene, partial sequence
JF919918.1	<i>Klebsiella pneumoniae</i> strain 15-10-2-9 16S ribosomal RNA gene, partial sequence
JF919853.1	<i>Acinetobacter baumannii</i> strain 7-10-9 16S ribosomal RNA gene, partial sequence
JF919917.1	<i>Klebsiella pneumoniae</i> strain 8-10-1-9 16S ribosomal RNA gene, partial sequence
JF919976.1	<i>Pseudomonas aeruginosa</i> strain 8-10-2-9 16S ribosomal RNA gene, partial sequence
JF919916.1	<i>Klebsiella pneumoniae</i> strain 41-9-9 16S ribosomal RNA gene, partial sequence
JF919881.1	<i>Escherichia coli</i> strain 20-10-1-9 16S ribosomal RNA gene, partial sequence
JF919919.1	<i>Klebsiella pneumoniae</i> strain 20-10-2-9 16S ribosomal RNA gene, partial sequence
JF919882.1	<i>Escherichia coli</i> strain 33-10-9 16S ribosomal RNA gene, partial sequence
JF920005.1	<i>Staphylococcus epidermidis</i> strain 13-10-9 16S ribosomal RNA gene, partial sequence
JF919894.1	<i>Enterobacter cancerogenus</i> strain 9-10-9 16S ribosomal RNA gene, partial sequence
JF919895.1	<i>Enterobacter cancerogenus</i> strain 23-10-9 16S ribosomal RNA gene, partial sequence
JF919998.1	<i>Staphylococcus aureus</i> strain 2-11-9 16S ribosomal RNA gene, partial sequence
JF919883.1	<i>Enterobacter cloacae</i> strain 1-11-1-9 16S ribosomal RNA gene, partial sequence
JF919925.1	<i>Klebsiella pneumoniae</i> strain 3-11-1-9 16S ribosomal RNA gene, partial sequence
JF920021.1	<i>Enterobacter</i> sp. 3-11-2-9 16S ribosomal RNA gene, partial sequence
JF919999.1	<i>Staphylococcus aureus</i> strain 3-11-3-9 16S ribosomal RNA gene, partial sequence
JF919857.1	<i>Acinetobacter baumannii</i> strain 28-10-9 16S ribosomal RNA gene, partial sequence

JF920020.1	Enterobacter sp. 29-10-2-9 16S ribosomal RNA gene, partial sequence
JF919856.1	Acinetobacter baumannii strain 27-10-9 16S ribosomal RNA gene, partial sequence
JF919891.1	Enterobacter hormaechei strain 8-11-1-9 16S ribosomal RNA gene, partial sequence
JF919926.1	Klebsiella pneumoniae strain 8-11-2-9 16S ribosomal RNA gene, partial sequence
JF919886.1	Enterococcus faecalis strain 12-11-1-9 16S ribosomal RNA gene, partial sequence
JF919892.1	Enterobacter hormaechei strain 12-11-2-9 16S ribosomal RNA gene, partial sequence
JF919980.1	Pseudomonas aeruginosa strain 10-12-1-9 16S ribosomal RNA gene, partial sequence
JF919860.1	Acinetobacter baumannii strain 10-12-2-9 16S ribosomal RNA gene, partial sequence
JF919858.1	Acinetobacter baumannii strain 4-12-9 16S ribosomal RNA gene, partial sequence
JF919859.1	Acinetobacter baumannii strain 8-12-9 16S ribosomal RNA gene, partial sequence
JF920000.1	Staphylococcus aureus strain 2-12-1-9 16S ribosomal RNA gene, partial sequence
JF920022.1	Enterobacter sp. 2-12-2-9 16S ribosomal RNA gene, partial sequence
JF919863.1	Acinetobacter baumannii strain 4-1-1-10 16S ribosomal RNA gene, partial sequence
JF919927.1	Klebsiella pneumoniae strain 4-1-2-10 16S ribosomal RNA gene, partial sequence
JF919864.1	Acinetobacter baumannii strain 7-1-1-10 16S ribosomal RNA gene, partial sequence
JF919928.1	Klebsiella pneumoniae strain 7-1-2-10 16S ribosomal RNA gene, partial sequence
JF920001.1	Staphylococcus aureus strain 9-2-10 16S ribosomal RNA gene, partial sequence
JF920009.1	Staphylococcus haemolyticus strain 5-12-9 16S ribosomal RNA gene, partial sequence
JF919979.1	Pseudomonas aeruginosa strain 3-12-9 16S ribosomal RNA gene, partial sequence
JF919868.1	Acinetobacter radioresistens strain 9-12-9 16S ribosomal RNA gene, partial sequence
JF919981.1	Pseudomonas aeruginosa strain 3-2-10 16S ribosomal RNA gene, partial sequence
JF919861.1	Acinetobacter baumannii strain 5-1-10 16S ribosomal RNA gene, partial sequence
JF919862.1	Acinetobacter baumannii strain 8-1-10 16S ribosomal RNA gene, partial sequence
JF919865.1	Acinetobacter baumannii strain 7-2-10 16S ribosomal RNA gene, partial sequence
JF920002.1	Staphylococcus aureus strain 2-3-1-10 16S ribosomal RNA gene, partial sequence
JF919887.1	Enterococcus faecalis strain 8-2-1-10 16S ribosomal RNA gene, partial sequence
JF920011.1	Staphylococcus hominis strain 5-2-10 16S ribosomal RNA gene, partial sequence
JF919874.1	Corynebacterium striatum strain 6-2-10 16S ribosomal RNA gene, partial sequence
JF919879.1	Escherichia coli strain 1-2-10 16S ribosomal RNA gene, partial sequence
JF919866.1	Acinetobacter baumannii strain 2-2-1-10 16S ribosomal RNA gene, partial sequence
JF919867.1	Acinetobacter baumannii strain 1-3-10 16S ribosomal RNA gene, partial sequence
JF919845.1	Acinetobacter baumannii strain 8-6-9 16S ribosomal RNA gene, partial sequence
JF919986.1	Staphylococcus aureus strain 7-3-8 16S ribosomal RNA gene, partial sequence
JF919889.1	Enterobacter hormaechei strain 4-3-1-8 16S ribosomal RNA gene, partial sequence
JF919884.1	Enterococcus faecalis strain 4-3-2-8 16S ribosomal RNA gene, partial sequence
JF919830.1	Acinetobacter baumannii strain 5-3-8 16S ribosomal RNA gene, partial sequence
JF919929.1	Morganella morganii strain 2-3-8 16S ribosomal RNA gene, partial sequence
JF919936.1	Pseudomonas aeruginosa strain 3-3-2-8 16S ribosomal RNA gene, partial sequence
JF919898.1	Klebsiella pneumoniae strain 2-4-8 16S ribosomal RNA gene, partial sequence
JF919988.1	Staphylococcus aureus strain 6-7-8 16S ribosomal RNA gene, partial sequence
JF919831.1	Acinetobacter baumannii strain 8-6-8 16S ribosomal RNA gene, partial sequence
JF919931.1	Morganella morganii strain 3-9-8 16S ribosomal RNA gene, partial sequence
JF919903.1	Klebsiella pneumoniae strain 4-9-8 16S ribosomal RNA gene, partial sequence
JF919938.1	Pseudomonas aeruginosa strain 5-8-8 16S ribosomal RNA gene, partial sequence
JF919942.1	Pseudomonas aeruginosa strain 5-10-8 16S ribosomal RNA gene, partial sequence

JF920026.1	Enterobacteriaceae bacterium 5-6-8 16S ribosomal RNA gene, partial sequence
JF919872.1	Corynebacterium striatum strain 6-8-8 16S ribosomal RNA gene, partial sequence
JF919896.1	Klebsiella granulomatis strain 2-7-8 16S ribosomal RNA gene, partial sequence
JF919987.1	Staphylococcus aureus strain 1-7-8 16S ribosomal RNA gene, partial sequence
JF920003.1	Staphylococcus epidermidis strain 3-7-8 16S ribosomal RNA gene, partial sequence
JF920023.1	Enterobacter sp. 8-8-8 16S ribosomal RNA gene, partial sequence
JF919939.1	Pseudomonas aeruginosa strain 15-8-8 16S ribosomal RNA gene, partial sequence
JF919875.1	Escherichia coli strain 11-8-8 16S ribosomal RNA gene, partial sequence
JF919900.1	Klebsiella pneumoniae strain 4-8-8 16S ribosomal RNA gene, partial sequence
JF919899.1	Klebsiella pneumoniae strain 3-6-8 16S ribosomal RNA gene, partial sequence
JF919941.1	Pseudomonas aeruginosa strain 1-10-8 16S ribosomal RNA gene, partial sequence
JF919834.1	Acinetobacter baumannii strain 14-8-8 16S ribosomal RNA gene, partial sequence
JF919945.1	Pseudomonas aeruginosa strain 14-10-8 16S ribosomal RNA gene, partial sequence
JF919922.1	Klebsiella pneumoniae strain 16-9-8 16S ribosomal RNA gene, partial sequence
JF920012.1	Serratia marcescens strain 27-10-8 16S ribosomal RNA gene, partial sequence
JF920013.1	Serratia marcescens strain 31-10-8 16S ribosomal RNA gene, partial sequence
JF919877.1	Escherichia coli strain 32-10-8 16S ribosomal RNA gene, partial sequence
JF919904.1	Klebsiella pneumoniae strain 5-9-8 16S ribosomal RNA gene, partial sequence
JF920016.1	Enterobacter sp. 6-9-8 16S ribosomal RNA gene, partial sequence
JF919989.1	Staphylococcus aureus strain 17-9-8 16S ribosomal RNA gene, partial sequence
JF920010.1	Staphylococcus hominis strain 9-8-8 16S ribosomal RNA gene, partial sequence
JF919930.1	Morganella morganii strain 2-9-8 16S ribosomal RNA gene, partial sequence
JF919901.1	Klebsiella pneumoniae strain 13-8-8 16S ribosomal RNA gene, partial sequence
JF919902.1	Klebsiella pneumoniae strain 16-8-8 16S ribosomal RNA gene, partial sequence
JF919835.1	Acinetobacter baumannii strain 2-10-8 16S ribosomal RNA gene, partial sequence
JF919894.1	Enterobacter cancerogenus strain 9-10-9 16S ribosomal RNA gene, partial sequence
JF919948.1	Pseudomonas aeruginosa strain 26-10-8 16S ribosomal RNA gene, partial sequence
JF920024.1	Enterobacter sp. 8-9-8 16S ribosomal RNA gene, partial sequence
JF919940.1	Pseudomonas aeruginosa strain 7-9-8 16S ribosomal RNA gene, partial sequence
JF920004.1	Staphylococcus epidermidis strain 12-10-8 16S ribosomal RNA gene, partial sequence
JF919933.1	Proteus mirabilis strain 5-11-1-8 16S ribosomal RNA gene, partial sequence
JF920028.1	Enterobacteriaceae bacterium 20-10-8 16S ribosomal RNA gene, partial sequence
JF920015.1	Stenotrophomonas maltophilia strain 21-10-8 16S ribosomal RNA gene, partial sequence
JF919952.1	Pseudomonas aeruginosa strain 34-10-8 16S ribosomal RNA gene, partial sequence
JF919943.1	Pseudomonas aeruginosa strain 6-10-8 16S ribosomal RNA gene, partial sequence
JF919990.1	Staphylococcus aureus strain 8-10-8 16S ribosomal RNA gene, partial sequence
JF919880.1	Escherichia coli strain 18-10-8 16S ribosomal RNA gene, partial sequence
JF919876.1	Escherichia coli strain 19-10-8 16S ribosomal RNA gene, partial sequence
JF919906.1	Klebsiella pneumoniae strain 25-10-8 16S ribosomal RNA gene, partial sequence
JF919949.1	Pseudomonas aeruginosa strain 28-10-8 16S ribosomal RNA gene, partial sequence
JF919836.1	Acinetobacter baumannii strain 17-10-8 16S ribosomal RNA gene, partial sequence
JF919951.1	Pseudomonas aeruginosa strain 33-10-8 16S ribosomal RNA gene, partial sequence
JF919956.1	Pseudomonas aeruginosa strain 9-11-8 16S ribosomal RNA gene, partial sequence
JF919962.1	Pseudomonas aeruginosa strain 10-5-9 16S ribosomal RNA gene, partial sequence

JF919964.1	<i>Pseudomonas aeruginosa</i> strain 12-5/9 16S ribosomal RNA gene, partial sequence
JF920017.1	<i>Enterobacter</i> sp. 3-6-1-9 16S ribosomal RNA gene, partial sequence
JF919885.1	<i>Enterococcus faecalis</i> strain 3-6-2-9 16S ribosomal RNA gene, partial sequence
JF919847.1	<i>Acinetobacter baumannii</i> strain 10-6-1-9 16S ribosomal RNA gene, partial sequence
JF920018.1	<i>Enterobacter</i> sp. 10-6-2-9 16S ribosomal RNA gene, partial sequence
JF919914.1	<i>Klebsiella pneumoniae</i> strain 26-9-9 16S ribosomal RNA gene, partial sequence
JF919975.1	<i>Pseudomonas aeruginosa</i> strain 27-9-9 16S ribosomal RNA gene, partial sequence
JF919923.1	<i>Klebsiella pneumoniae</i> strain 7-9-9 16S ribosomal RNA gene, partial sequence
JF919921.1	<i>Klebsiella pneumoniae</i> strain 10-3-10 16S ribosomal RNA gene, partial sequence
JF919910.1	<i>Klebsiella pneumoniae</i> strain 17-9-1-9 16S ribosomal RNA gene, partial sequence
JF919973.1	<i>Pseudomonas aeruginosa</i> strain 17-9-2-9 16S ribosomal RNA gene, partial sequence
JF919915.1	<i>Klebsiella pneumoniae</i> strain 30-9-1-9 16S ribosomal RNA gene, partial sequence
JF920019.1	<i>Enterobacter</i> sp. 30-9-2-9 16S ribosomal RNA gene, partial sequence
JF919888.1	<i>Enterococcus gallinarum</i> strain 8-9-9 16S ribosomal RNA gene, partial sequence
JF919851.1	<i>Acinetobacter baumannii</i> strain 9-9-1-9 16S ribosomal RNA gene, partial sequence
JF919970.1	<i>Pseudomonas aeruginosa</i> strain 9-9-2-9 16S ribosomal RNA gene, partial sequence
JF919852.1	<i>Acinetobacter baumannii</i> strain 10-9-1-9 16S ribosomal RNA gene, partial sequence
JF919971.1	<i>Pseudomonas aeruginosa</i> strain 10-9-2-9 16S ribosomal RNA gene, partial sequence
JF919972.1	<i>Pseudomonas aeruginosa</i> strain 16-9-9 16S ribosomal RNA gene, partial sequence
JF919977.1	<i>Pseudomonas aeruginosa</i> strain 16-10-9 16S ribosomal RNA gene, partial sequence
JF919854.1	<i>Acinetobacter baumannii</i> strain 10-10-9 16S ribosomal RNA gene, partial sequence
JF919855.1	<i>Acinetobacter baumannii</i> strain 22-10-1-9 16S ribosomal RNA gene, partial sequence
JF919935.1	<i>Proteus mirabilis</i> strain 22-10-2-9 16S ribosomal RNA gene, partial sequence
JF919924.1	<i>Klebsiella pneumoniae</i> strain 25-10-1-9 16S ribosomal RNA gene, partial sequence
JF919978.1	<i>Pseudomonas aeruginosa</i> strain 25-10-2-9 16S ribosomal RNA gene, partial sequence

4 Apêndice - Resultado do alinhamento contra os bancos de dados de referência

Apêndice – Resultado dos alinhamentos realizados para as 200 sequências identificadas, para os bancos de dados GreenGenes, SILVA e RDP-II, bem como o índice de similaridade e o desvio padrão (SD) entre os bancos de dados.

id GenBank	Espécies Identificadas	%Greengenes	% SILVA	% RDP-2	SD
JF919830.1	Acinetobacter baumannii	99,13	99,13	99,13	0,000
JF919831.1	Acinetobacter baumannii	99,55	99,55	99,55	0,000
JF919832.1	Acinetobacter baumannii	99,78	99,78	99,78	0,000
JF919833.1	Acinetobacter baumannii	100,00	100,00	100,00	0,000
JF919834.1	Acinetobacter baumannii	99,78	99,79	99,79	0,006
JF919835.1	Acinetobacter baumannii	100,00	100,00	100,00	0,000
JF919836.1	Acinetobacter baumannii	99,74	99,74	99,74	0,000
JF919837.1	Acinetobacter baumannii	99,55	99,55	99,55	0,000
JF919838.1	Acinetobacter baumannii	99,58	99,59	99,59	0,006
JF919839.1	Acinetobacter baumannii	100,00	100,00	100,00	0,000
JF919840.1	Acinetobacter baumannii	99,57	99,57	99,57	0,000
JF919841.1	Acinetobacter baumannii	99,75	99,75	99,75	0,000
JF919842.1	Acinetobacter baumannii	99,43	99,43	99,72	0,167
JF919843.1	Acinetobacter baumannii	99,79	99,80	99,80	0,006
JF919844.1	Acinetobacter baumannii	99,76	99,76	99,76	0,000
JF919845.1	Acinetobacter baumannii	99,79	99,59	99,59	0,115
JF919846.1	Acinetobacter baumannii	99,77	99,77	99,77	0,000
JF919847.1	Acinetobacter baumannii	99,57	99,57	99,78	0,121
JF919848.1	Acinetobacter baumannii	99,16	99,16	99,16	0,000
JF919849.1	Acinetobacter baumannii	99,58	99,39	99,39	0,110
JF919850.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919851.1	Acinetobacter baumannii	100,00	100,00	100,00	0,000
JF919852.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919853.1	Acinetobacter baumannii	99,08	99,08	99,08	0,000
JF919854.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919855.1	Acinetobacter baumannii	99,58	99,79	99,79	0,121
JF919856.1	Acinetobacter baumannii	99,33	99,35	99,35	0,012
JF919857.1	Acinetobacter baumannii	99,73	99,73	99,73	0,000
JF919858.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919859.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919860.1	Acinetobacter baumannii	99,79	99,60	99,60	0,110
JF919861.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919862.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919863.1	Acinetobacter baumannii	99,78	99,78	99,78	0,000
JF919864.1	Acinetobacter baumannii	99,78	99,78	99,78	0,000
JF919865.1	Acinetobacter baumannii	99,79	99,80	99,80	0,006
JF919866.1	Acinetobacter baumannii	99,79	99,79	99,79	0,000
JF919867.1	Acinetobacter baumannii	99,73	99,73	99,73	0,000
JF919868.1	Acinetobacter radioresistens	99,36	99,57	99,57	0,121
JF919869.1	Bacillus clausii	99,76	99,76	99,76	0,000
JF919870.1	Brevibacillus agri	99,37	99,37	99,37	0,000
JF919872.1	Corynebacterium striatum	99,78	99,56	99,78	0,127
JF919873.1	Corynebacterium striatum	100,00	100,00	100,00	0,000
JF919874.1	Corynebacterium striatum	100,00	100,00	99,65	0,202
JF919874.1	Corynebacterium striatum	100,00	100,00	100,00	0,000

JF919894.1	Enterobacter cancerogenus	98,75	98,75	98,75	0,000
JF919894.1	Enterobacter cancerogenus	99,77	99,77	99,77	0,000
JF919895.1	Enterobacter cancerogenus	96,45	96,45	96,45	0,000
JF919883.1	Enterobacter cloacae	100,00	100,00	100,00	0,000
JF920016.1	Enterobacter hormaechei	99,56	99,56	99,78	0,127
JF920017.1	Enterobacter hormaechei	99,33	99,33	99,33	0,000
JF920018.1	Enterobacter hormaechei	99,36	99,36	99,58	0,127
JF920019.1	Enterobacter hormaechei	99,79	99,80	99,80	0,006
JF920020.1	Enterobacter hormaechei	99,58	99,59	99,59	0,006
JF920021.1	Enterobacter hormaechei	99,79	99,80	99,80	0,006
JF920022.1	Enterobacter hormaechei	99,58	99,58	99,79	0,121
JF920023.1	Enterobacter hormaechei	99,58	99,58	99,79	0,121
JF920024.1	Enterobacter hormaechei	98,23	98,23	98,67	0,254
JF919889.1	Enterobacter hormaechei	99,58	99,59	99,79	0,118
JF919890.1	Enterobacter hormaechei	100,00	100,00	100,00	0,000
JF919891.1	Enterobacter hormaechei	99,79	99,79	100,00	0,121
JF919892.1	Enterobacter hormaechei	99,30	99,30	99,30	0,000
JF919893.1	Enterobacter hormaechei	99,77	99,55	99,55	0,127
JF919884.1	Enterococcus faecalis	100,00	100,00	100,00	0,000
JF919885.1	Enterococcus faecalis	100,00	100,00	100,00	0,000
JF919886.1	Enterococcus faecalis	100,00	100,00	100,00	0,000
JF919887.1	Enterococcus faecalis	99,79	99,40	99,60	0,195
JF919888.1	Enterococcus gallinarum	100,00	100,00	100,00	0,000
JF920026.1	Escherichia coli	99,19	99,19	99,19	0,000
JF920028.1	Escherichia coli	100,00	99,76	100,00	0,139
JF920029.1	Escherichia coli	98,03	98,03	98,36	0,191
JF919875.1	Escherichia coli	99,37	99,16	99,57	0,205
JF919876.1	Escherichia coli	99,76	99,51	99,76	0,144
JF919877.1	Escherichia coli	99,52	99,28	99,52	0,139
JF919878.1	Escherichia coli	99,52	99,29	99,52	0,133
JF919879.1	Escherichia coli	99,58	99,39	99,79	0,200
JF919880.1	Escherichia coli	99,49	99,49	99,49	0,000
JF919881.1	Escherichia coli	99,79	99,79	99,79	0,000
JF919881.1	Escherichia coli	99,53	99,53	99,53	0,000
JF919882.1	Escherichia coli	100,00	100,00	99,77	0,133
JF919896.1	Klebsiella granulomatis	100,00	100,00	100,00	0,000
JF919897.1	Klebsiella granulomatis	100,00	100,00	100,00	0,000
JF920025.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919898.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919899.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919900.1	Klebsiella pneumoniae	99,74	99,74	100,00	0,150
JF919901.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919902.1	Klebsiella pneumoniae	99,58	99,59	99,59	0,006
JF919903.1	Klebsiella pneumoniae	99,26	99,26	99,26	0,000
JF919904.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919905.1	Klebsiella pneumoniae	99,76	99,76	99,76	0,000
JF919906.1	Klebsiella pneumoniae	100,00	99,34	99,56	0,336
JF919907.1	Klebsiella pneumoniae	99,79	99,79	99,79	0,000
JF919908.1	Klebsiella pneumoniae	99,79	99,39	99,39	0,231

JF919909.1	Klebsiella pneumoniae	99,16	99,16	99,16	0,000
JF919910.1	Klebsiella pneumoniae	99,59	99,00	99,00	0,341
JF919911.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919913.1	Klebsiella pneumoniae	99,55	99,33	99,55	0,127
JF919914.1	Klebsiella pneumoniae	99,37	99,19	99,19	0,104
JF919915.1	Klebsiella pneumoniae	99,79	99,59	100,00	0,205
JF919916.1	Klebsiella pneumoniae	99,79	99,40	99,60	0,195
JF919917.1	Klebsiella pneumoniae	99,79	99,59	99,59	0,115
JF919918.1	Klebsiella pneumoniae	99,16	99,17	99,17	0,006
JF919919.1	Klebsiella pneumoniae	99,37	99,38	99,38	0,006
JF919919.1	Klebsiella pneumoniae	99,79	99,80	99,80	0,006
JF919921.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919922.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919923.1	Klebsiella pneumoniae	100,00	100,00	100,00	0,000
JF919924.1	Klebsiella pneumoniae	99,58	99,58	99,58	0,000
JF919925.1	Klebsiella pneumoniae	99,58	99,58	99,58	0,000
JF919926.1	Klebsiella pneumoniae	99,58	99,00	99,00	0,335
JF919927.1	Klebsiella pneumoniae	99,58	99,58	99,58	0,000
JF919928.1	Klebsiella pneumoniae	98,85	98,85	98,85	0,000
JF919929.1	Morganella morganii	100,00	100,00	100,00	0,000
JF919930.1	Morganella morganii	100,00	100,00	100,00	0,000
JF919931.1	Morganella morganii	100,00	100,00	100,00	0,000
JF919932.1	Proteus mirabilis	100,00	100,00	100,00	0,000
JF919933.1	Proteus mirabilis	99,79	99,79	99,79	0,000
JF919934.1	Proteus mirabilis	99,12	99,12	99,34	0,127
JF919935.1	Proteus mirabilis	100,00	100,00	100,00	0,000
JF919984.1	Providencia stuartii	99,57	99,57	99,57	0,000
JF919985.1	Providencia stuartii	99,32	99,32	99,32	0,000
JF919936.1	Pseudomonas aeruginosa	98,47	98,47	99,11	0,370
JF919937.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919938.1	Pseudomonas aeruginosa	99,76	99,76	99,76	0,000
JF919939.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919940.1	Pseudomonas aeruginosa	99,57	99,57	99,79	0,127
JF919941.1	Pseudomonas aeruginosa	99,48	99,48	99,48	0,000
JF919942.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919943.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919945.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919948.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919949.1	Pseudomonas aeruginosa	99,33	99,33	99,55	0,127
JF919950.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919951.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919952.1	Pseudomonas aeruginosa	99,77	99,78	99,78	0,006
JF919953.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919954.1	Pseudomonas aeruginosa	99,73	99,73	99,73	0,000
JF919955.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919956.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919957.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919958.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919959.1	Pseudomonas aeruginosa	99,58	99,58	99,79	0,121

JF919960.1	Pseudomonas aeruginosa	99,79	99,38	99,38	0,237
JF919961.1	Pseudomonas aeruginosa	99,53	99,53	99,53	0,000
JF919962.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919963.1	Pseudomonas aeruginosa	99,37	99,18	99,18	0,110
JF919964.1	Pseudomonas aeruginosa	99,53	99,53	99,53	0,000
JF919965.1	Pseudomonas aeruginosa	99,50	99,50	99,75	0,144
JF919966.1	Pseudomonas aeruginosa	99,79	99,79	100,00	0,121
JF919967.1	Pseudomonas aeruginosa	99,79	99,39	99,39	0,231
JF919968.1	Pseudomonas aeruginosa	99,79	99,79	99,79	0,000
JF919969.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919970.1	Pseudomonas aeruginosa	99,58	99,39	100,00	0,312
JF919971.1	Pseudomonas aeruginosa	99,79	99,79	99,79	0,000
JF919972.1	Pseudomonas aeruginosa	99,37	99,38	99,38	0,006
JF919973.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919975.1	Pseudomonas aeruginosa	100,00	100,00	100,00	0,000
JF919976.1	Pseudomonas aeruginosa	99,58	99,37	99,58	0,121
JF919977.1	Pseudomonas aeruginosa	99,79	99,79	99,79	0,000
JF919978.1	Pseudomonas aeruginosa	99,55	99,55	99,77	0,127
JF919979.1	Pseudomonas aeruginosa	99,79	99,79	100,00	0,121
JF919980.1	Pseudomonas aeruginosa	100,00	100,00	99,77	0,133
JF919981.1	Pseudomonas aeruginosa	99,76	99,76	99,31	0,260
JF919982.1	Pseudomonas cedrina	99,79	99,79	99,79	0,000
JF919982.1	Pseudomonas cedrina	99,55	99,55	99,77	0,127
JF919983.1	Pseudomonas putida	99,79	99,39	99,39	0,231
JF919983.1	Pseudomonas putida	99,36	99,36	99,57	0,121
JF920006.1	Serratia marcescens	99,79	99,58	99,79	0,121
JF920007.1	Serratia marcescens	100,00	100,00	100,00	0,000
JF920012.1	Serratia marcescens	100,00	100,00	100,00	0,000
JF920013.1	Serratia marcescens	99,79	99,60	99,60	0,110
JF919986.1	Staphylococcus aureus	99,78	99,78	99,78	0,000
JF919987.1	Staphylococcus aureus	99,59	99,60	100,00	0,234
JF919988.1	Staphylococcus aureus	99,54	99,54	99,54	0,000
JF919989.1	Staphylococcus aureus	100,00	100,00	100,00	0,000
JF919989.1	Staphylococcus aureus	99,79	99,60	99,60	0,110
JF919990.1	Staphylococcus aureus	99,66	99,66	99,66	0,000
JF919991.1	Staphylococcus aureus	100,00	100,00	100,00	0,000
JF919992.1	Staphylococcus aureus	100,00	100,00	100,00	0,000
JF919993.1	Staphylococcus aureus	100,00	100,00	100,00	0,000
JF919994.1	Staphylococcus aureus	99,80	99,80	99,80	0,000
JF919995.1	Staphylococcus aureus	99,79	99,80	99,80	0,006
JF919996.1	Staphylococcus aureus	99,80	99,80	99,80	0,000
JF919997.1	Staphylococcus aureus	100,00	100,00	100,00	0,000
JF919998.1	Staphylococcus aureus	99,79	99,79	100,00	0,121
JF919999.1	Staphylococcus aureus	99,79	99,79	100,00	0,121
JF920000.1	Staphylococcus aureus	99,78	99,78	99,78	0,000
JF920001.1	Staphylococcus aureus	99,79	99,80	99,80	0,006
JF920002.1	Staphylococcus aureus	99,79	99,60	99,60	0,110
JF920003.1	Staphylococcus epidermidis	100,00	100,00	100,00	0,000
JF920004.1	Staphylococcus epidermidis	100,00	100,00	100,00	0,000

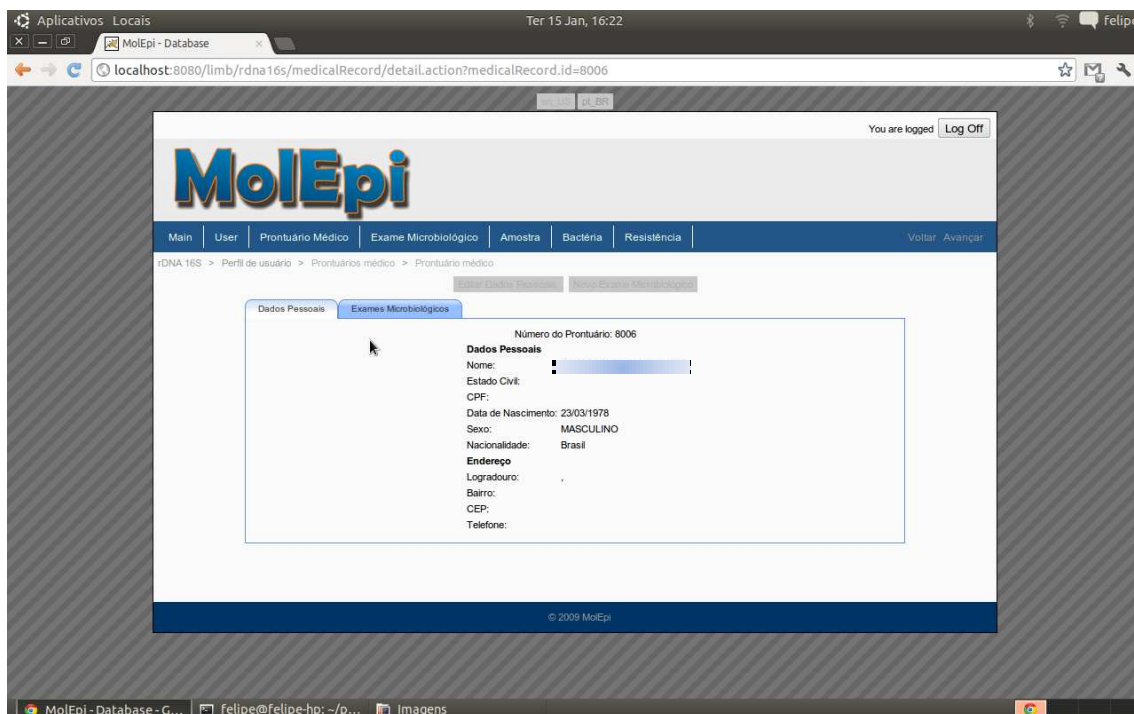
JF920005.1	Staphylococcus epidermidis	100,00	100,00	100,00	0,000
JF920005.1	Staphylococcus epidermidis	99,58	99,59	99,79	0,118
JF920008.1	Staphylococcus haemolyticus	99,76	99,76	99,76	0,000
JF920009.1	Staphylococcus haemolyticus	100,00	100,00	100,00	0,000
JF920010.1	Staphylococcus hominis	99,59	99,59	99,59	0,000
JF920011.1	Staphylococcus hominis	100,00	100,00	100,00	0,000
JF920014.1	Staphylococcus sciuri	100,00	100,00	100,00	0,000
JF920015.1	Stenotrophomonas maltophilia	100,00	100,00	100,00	0,000

5 Apêndice - Interface das principais aplicações do MolEpi

Apêndice – Visualização das telas do sistema MolEpi.



01 – Tela inicial do MolEpi.



02 – Informações básicas do paciente, inseridas no prontuário eletrônico.

Dados Pessoais

Exames Microbiológicos

+ Exame Microbiológico: **15**, em 22/02/2010 06:31:04
+ Exame Microbiológico: **21**, em 22/02/2010 07:38:48
+ Exame Microbiológico: **22**, em 22/02/2010 07:54:43
+ Exame Microbiológico: **79**, em 29/03/2010 04:24:07
+ Exame Microbiológico: **83**, em 05/04/2010 09:50:34

Editar Exame Microbiológico

Nova Amostra

Dado epidemiológico:
Possível Fonte de Infecção:
Setor de Coleta: Enfermaria
Procedimento Invasivo: Traqueostomia
Data: 05/04/2010 09:50:34

Amostra: **74**

Editar Amostra

Nova Bactéria

Tipo de Amostra: Cultura trato respiratorio
Subtipo de amostra: aspirado traqueal

Bactéria: **72**

Editar Bactéria

Novo Diagnóstico Molecular

Nova Resistência

Descrição: JF919989.1
Diagnóstico Laboratorial: Pseudomonas aeruginosa

Diagnóstico Molecular

Database	Description	Score	E Value	Percentage Identity
GenBank	Pseudomonas aeruginosa strain 1248 16S ribosomal RNA gene, partial sequence	790.0		
GenBank	Pseudomonas aeruginosa strain TMSS 16S ribosomal RNA gene, partial sequence	790.0	0.0	100.0

Resistência: 71

Editar Teste de Resistência

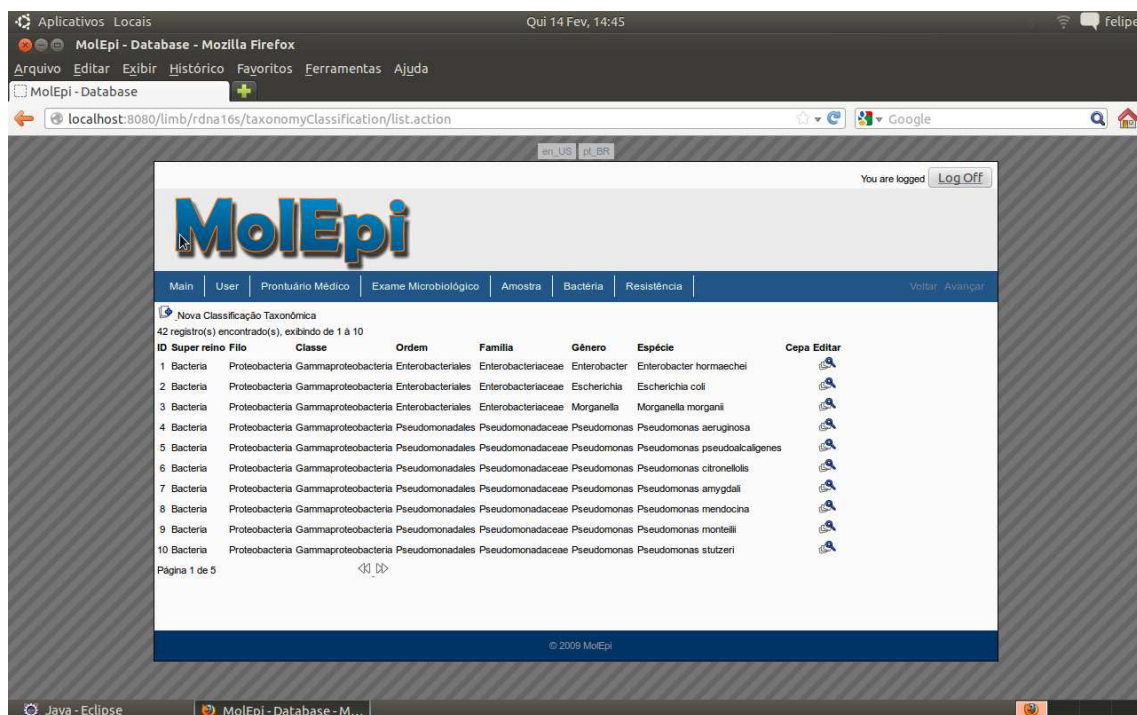
Novo Teste de Antibiótico

Laboratório: LABMEG - Campus Arapiraca

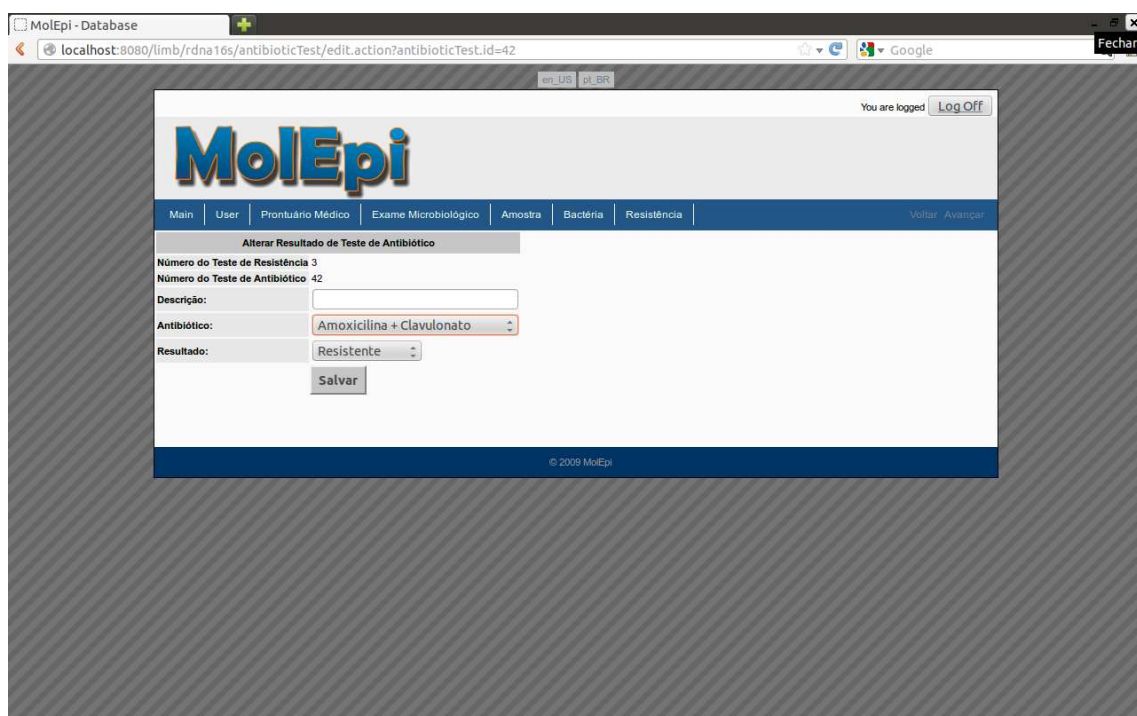
Antibiograma

Ceftazidima Resistente	Editar	Remover
Amoxicilina + Clavulonato Resistente	Editar	Remover
Aztreonam Resistente	Editar	Remover
Cefoxitina Resistente	Editar	Remover
Amoxicilina Resistente	Editar	Remover
Ceftriaxona Resistente	Editar	Remover
Cloranfenicol Resistente	Editar	Remover
Ciprofloxacina Resistente	Editar	Remover
Cefepime Sensível	Editar	Remover

03 – Prontuário eletrônico do paciente com dados do processo infeccioso, identificação molecular e teste de resistência.



04 – Taxonomia utilizada para vincular diagnóstico dos microrganismos.



05 – Cadastro do resultado do teste de resistência a antimicrobianos.

Novo Diagnóstico Molecular

Número da Bactéria: 1

Descrição:

Sequência:

```

CTTCTTCTGTCGGTACGTCATTGCCAAGGTTATTAACCTTGACCCTTCCTCCCGACTGAAAG
TCTTTACAACCCGAAGGCCTTCTTCATACACGCGGCATGGCTGCATCAGGCTTGCGCCATT
GTGCAATATTTCCCACTGCTGCCTCCCGTAGGAGTCTGGGCCGTGTCTCAGTCCCAGTGTGG
CTGATCATCCTCTCAGACCAGCTAGGGATCGTCGCCTAGGTAAGCCGTTACCTCACCTACTA
GCTAATCCCATATGGGTTTCATCTGATGGCGCGAGGCCCGAAGGTCCCCCGCTTTGGTCCGAA
GACATTATGCGGTATTAGCTACCGTTTCCAGTAGTTATCCCCGCCATCAGGCAGATCCCCA
TACATTACTACCCGTCCGCCGCTCGTCAGCAGAGAAGCAAGCTTCTCCCTGTTACCGCTCG
A

```

Banco de dados: genbank

genbank
greengene
local
rdp2
silva

Blast

06 – Identificação integradas aos principais bancos de dados de 16S rRNA.

Main	User	Prontuário Médico	Exame Microbiológico	Amostra	Bactéria	Resistência	Voltar	Avançar
rDNA 16S > Perfil de usuário > Testes de Resistência								
Resultado (BLASTN), em rdp2								
Selecionar	ID	Descrição	Score	EValue				
☐	Id S001549854	Morganella morganii; ATCC 8076; FJ971868	399.0	0.0				
☐	Id S001352604	Morganella morganii; HX081027; FJ858185	399.0	0.0				
☐	Id S000973565	uncultured Morganella sp.; 8817-D4-C-1F; EU289074	399.0	0.0				
☐	Id S000648024	Morganella morganii; PW2; DQ355171	399.0	0.0				
☐	Id S000410342	Morganella morganii; CPD30; AY464464	399.0	0.0				
☐	Id S001188984	Morganella morganii; BH-16; FJ219587	395.0	0.0				
☐	Id S000973586	uncultured Morganella sp.; 8817-D4-C-10A; EU289095	395.0	0.0				
☐	Id S000458072	Morganella morganii; N32; AB182194	395.0	0.0				
☐	Id S000458058	Morganella morganii; N17; AB182180	395.0	0.0				
☐	Id S000458057	Morganella morganii; N16; AB182179	395.0	0.0				
☐	Id S000388834	activated sludge bacterium BC55; AF189697	392.0	0.0				

07 – Resultado do alinhamento local utilizado na recuperação das sequências mais similares.

Home
Access MOLEPI DB
Relational Publications
Tools

Selecionar bactérias

• Os dados não preenchidos não serão incluídos nas condições da consulta

Consulta

Medical Record

Sexo: MASCULINO

Exame Microbiológico

Setor de Coleta: Ambulatorio

Resistência

Antibiótico: Amicacina

Resultado: Intermediario

Taxonomia

Espécie: Pseudomonas aeruginosa

Buscar

Resultado

Programa: ClustalW

Sequência de referência: Pseudomonas aeruginosa -NBRAJG90/EU734822

Bactérias

Selecionar (Marcar tudo / Desmarcar tudo)	ID	Descrição	Sequência
☒	61	Pseudomonas aeruginosa	aggtattaac...
☒	3	Pseudomonas aeruginosa	GTCAAACAGC...
☒	13	Pseudomonas aeruginosa	cgaagacctt...
☒	47	Pseudomonas aeruginosa	cagcaggtat...
☒	44	Pseudomonas aeruginosa	taattacta...

08 – Interface de recuperação de sequência de acordo com parâmetros pré-definidos para estudo e mapeamento epidemiológico e filogenético.

Home
Access MOLEPI DB
Relational Publications
Tools

Resultado

View

Jalview Start Jalview

Download

[.aln http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.aln](http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.aln)
[.dnd http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.dnd](http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.dnd)
[.infile http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.infile](http://localhost:8080/lmb/rdna16s/msa/download.action?alignFile=53F6E6D8BA8A77B530DBDD97F57EFC6D.infile)

ALN

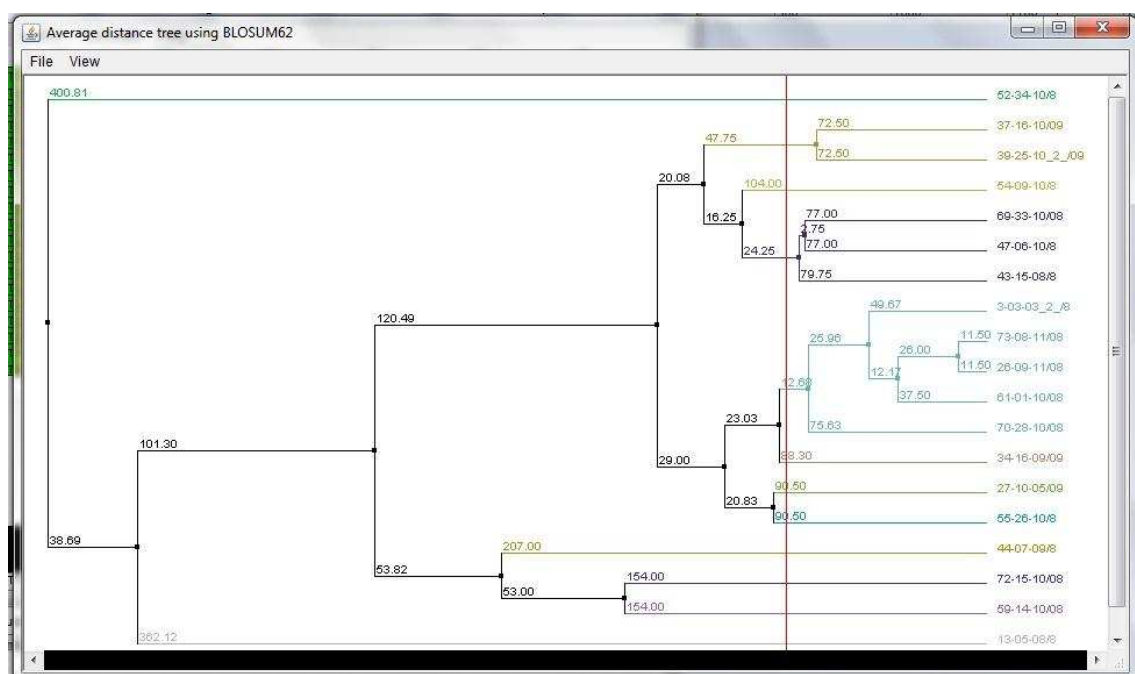
CLUSTAL 2.0.11 multiple sequence alignment

37-16-10/09 TGGTACGTCCAACAGCA-GGTATTAACTTACTGCCCTTCCTCCCAACTTAAAGTGCTTTA
39-25-10_2_/09 TGGTACGTCCCACAGCA-GGTATTAACTTACTGCCCTTCCTCCCAACTTAA-GTGCTTTA
52-34-10/8 -----TATTAACTTACTGCCCTTCCTCCCA-CTTAA-GTGCTTTA
54-09-10/8 -----ACAGCA-GGTATTAACTTACTGCCCTTCCTCCCAACTTAAAGTGCTTTA
69-33-10/08 -----ACAGCA-GGTATTAACTTACTGCCCTTCCTCCCAACTTAAAGTGCTTTA
44-07-09/8 -----TAACTTACTGCCCTTCCTCCCAACTTAAAGTGCTTTA

09 – Alinhamento Global com o ClustalW, para estudos filogenéticos.



10 – Visualização do alinhamento múltiplo através do editor de sequências JalView.



11 – Árvore filogenética construída a partir as sequências recuperadas do MolEpiDB.

1 Anexo - Paracer do Comitê de Ética



**UNIVERSIDADE FEDERAL DE ALAGOAS
COMITÊ DE ÉTICA EM PESQUISA**

Maceió – AL, 17/04/2007

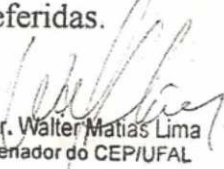
Senhor (a) Pesquisador (a), Tiago Gomes de Andrade
Marcelo Costa Oliveira
Janaina Ferro Pereira
Regina Célia Sales Santos
Cristiane Araújo Nascimento
Maria Betânia Monteiro de Farias
Ana Paula Nogueira de Magalhães

O Comitê de Ética em Pesquisa (CEP), reunido em 16/04/2007 e com base no parecer emitido pelo (a) relator (a) do processo nº 014852/2006-29 sob o título **Vigilância epidemiológica na Unidade de Emergência do Agreste: implantação de serviço de epidemiologia molecular e criação de banco de dados informatizado para seqüências de rDNA 16S de bactérias patogênicas**, de sua autoria, vem por meio deste instrumento comunicar sua aprovação com base no item VIII.13, b, da Resolução nº 196/96.

Outrossim, recomendamos a observância do que consta na folha de rosto com respeito ao cumprimento dos prazos para entrega de relatórios, bem como o atendimento da referida Resolução da CONEP/CNS, quando for o caso (*).

Na eventualidade de esclarecimentos adicionais, este Comitê coloca-se a disposição dos interessados para o acompanhamento da pesquisa em seus dilemas éticos e exigências contidas nas Resoluções supra referidas.

(*) Áreas temáticas especiais


Prof. Dr. Walter Matias Lima
Coordenador do CEPI/UFAL



Documento recebido

04 / 05 / 2007

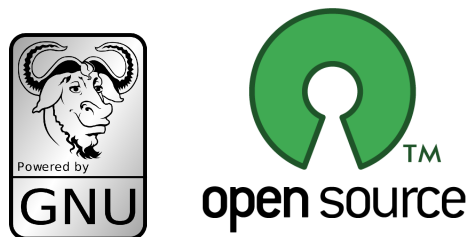


Assinatura do Funcionário

2 Anexo - Registros Instituto Nacional de Propriedade Industrial - INPI

A Licença GNU GPL (Licença Pública Geral), oferece ao desenvolvedor a possibilidade de lançar seus respectivos sistemas de softwares de maneira a não vetar a cópia, utilização, alteração e distribuição por qualquer outro que deseje manipular tal sistema. Como qualquer licença de software ela também exige a execução dos termos dispostos, a fim de assegurar os direitos do autor ou desenvolvedor e do usuário final.

A GPL versão 3, em termos gerais assegura que toda aplicação licenciada nos seus termos, deve ser distribuída por completo, ou seja, toda aplicação deve ser liberada como *Open Source*. Ainda, deve ser distribuído juntamente ou disposto em meios de acesso ao público (usuários), o código fonte da aplicação para cópia, uso, alteração e ou distribuição do mesmo, também nos termos da GPL.



Protocolo do registro no INPI do BACSearch

INPI

protocolo

31 AGO 11 28 2012 000194

PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR

IDENTIFICAÇÃO DO PEDIDO (Para uso do INPI)

Número do Pedido _____ Protocolo, Data e Hora _____

DADOS DO AUTOR DO PROGRAMA

Nº de Autores 5 Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF* 030.570.763-75

Nome FELIPE JOSÉ DE QUEIROZ SARMENTO

Nome Abreviado, pseudônimo ou sinal convencional (se houver) FELIPE SARMENTO

Data de Nascimento 22/12/1978 Nacionalidade BRASILEIRA

Endereço RESIDENCIAL VERT PARADISO, QD.E, LOT.04, Nº 37 ANTARES

Cidade MACEIO UF AL País BRASIL

CEP 57.048-375 Telefone 8230325311 FAX _____

E-mail felipesarmento@gmail.com

DADOS DO TITULAR DOS DIREITOS PATRIMONIAIS

Nº de Titulares 1 Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF/CNPJ* 24098477000110

Nome/Razão Social UNIVERSIDADE FEDERAL DA PARAÍBA

Nome abreviado, pseudônimo ou sinal convencional (se houver) UFPB

Data de Nascimento 13/12/1960 Nacionalidade/Origem BRASIL

Endereço CIDADE UNIVERSITÁRIA, CAMPUS I - REITORIA, 1ª ANDAR

Cidade JOÃO PESSOA UF PB País BRASIL

CEP 58.051-900 Telefone 8332167558 FAX _____

E-mail cit@prpg.ufpb.br

☒ **SIM**, este Titular é Pessoa Jurídica. Caso afirmativo, assinale a melhor classificação:

☐ Órgão Público ☐ Sociedade com Intuito não Econômico ☐ Microempresa ☐ Software House

☒ Instituição Pública de Ensino ou Pesquisa ☐ Instituição Privada de Ensino ou Pesquisa ☐ Outras

ENDEREÇO PARA CORRESPONDÊNCIA E CONTATO (Preencha apenas o necessário)

Toda correspondência será enviada para: ☐ O Procurador ou ☒ O Titular acima ou

☐ Escaninho nº _____ ☐ Representação INPI em: _____ ☒ O Endereço abaixo:

Nome FELIPE JOSÉ DE QUEIROZ SARMENTO

Endereço RESID. VERT PARADISO, QD.E LOT. 04, Nº 37 - ANTARES

Cidade MACEIO UF AL País BRASIL

CEP 57.048-375 Telefone 8230325311 FAX _____

E-mail felipesarmento@gmail

Modelo I (folha 1/2) E

Protocolo do registro no INPI do MolEpi

INPI

- 3 OUT 11 54 2012 000221

PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR

DADOS DO PEDIDO

Número do Pedido _____ Protocolo, Data e Hora _____

DADOS DO AUTOR

Nº de Autores **5** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF* **030.570.763-75**

Nome **FELIPE JOSÉ DE QUEIROZ SARMENTO**

Nome Abreviado, pseudônimo ou sinal convencional (se houver) **FELIPE SARMENTO**

Data de Nascimento **22/12/1978** Nacionalidade **BRASILEIRA**

Endereço **RESIDENCIAL VERT PARADISO, QD.E, LOT.04, Nº 37 ANTARES**

Cidade **MACEIO** UF **AL** País **BRASIL**

CEP **57.048-375** Telefone **8230325311** FAX _____

E-mail **felipesarmento@gmail.com**

DADOS DO TITULAR

Nº de Titulares **1** Se mais de um, preencha a "Continuação", com todos os dados solicitados neste Quadro. Date e assine.

CPF/CNPJ* **24098477000110**

Nome/Razão Social **UNIVERSIDADE FEDERAL DA PARAÍBA**

Nome abreviado, pseudônimo ou sinal convencional (se houver) **UFPB**

Data de Nascimento **13/12/1960** Nacionalidade/Origem **BRASIL**

Endereço **CIDADE UNIVERSITÁRIA, CAMPUS I - REITORIA, 1º ANDAR**

Cidade **JOÃO PESSOA** UF **PB** País **BRASIL**

CEP **58.051-900** Telefone **8332167558** FAX _____

E-mail **cit@prpg.ufpb.br**

☒ **SIM**, este Titular é Pessoa Jurídica. Caso afirmativo, assinale a melhor classificação:

☐ Órgão Público ☐ Sociedade com Intuito não Econômico ☐ Microempresa ☐ Software House

☒ Instituição Pública de Ensino ou Pesquisa ☐ Instituição Privada de Ensino ou Pesquisa ☐ Outras

ENDEREÇO PARA CORRESPONDÊNCIA

Toda correspondência será enviada para: ☐ O Procurador ou ☒ O Titular acima ou

☐ Escaninho nº _____ ☐ Representação INPI em: _____ ☒ O Endereço abaixo:

Nome **FELIPE JOSÉ DE QUEIROZ SARMENTO**

Endereço **RESID. VERT PARADISO, QD.E LOT. 04, Nº 37 - ANTARES**

Cidade **MACEIO** UF **AL** País **BRASIL**

CEP **57.048-375** Telefone **8230325311** FAX _____

E-mail **felipesarmento@gmail**

Modelo 1 (folha 1/2) E



Esta tese de doutorado, composta pelo autor em fonte Computer Modern, papel A4 em gramatura $90g/m^2$, foi editada utilizando o pacote COPPE_{TEX}, através do processador de texto L^AT_EX3e que possui licença GNU GPL3. As referências bibliográficas foram formatadas de acordo com a norma NBR6023:2002, e as citações às referências formatadas de acordo com a norma NBR10520:2002 com uso do pacote abnTeX2.

ABN_{TEX}2

João Pessoa, 16 setembro de 2013.