Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Matemática e Estatística

Bruna Garcia Cruz Ribeiro

Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo Laboratório de HLA da UERJ

Rio de Janeiro

2017

Bruna Garcia Cruz Ribeiro

Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo Laboratório de HLA da UERJ

Dissertação apresentada como requisito parcial para obtenção do título de Mestre ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Orientador: Prof. Dr. Dilson Silva

Coorientador: Prof. Dr. Alexandre da Costa Sena

Rio de Janeiro

2017

CATALOGAÇÃO NA FONTE
UERJ / REDE SIRIUS / BIBLIOTECA CTC-A

R484 Ribeiro, Bruna Garcia Cruz
Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo laboratório de HLA da UERJ/ Bruna Garcia Cruz Ribeiro. – 2017.
128f. : il.

Orientador: Dilson Silva
Coorientador: Alexandre da Costa Sena
Dissertação (Mestrado em Ciências Computacionais) - Universidade do Estado do Rio de Janeiro, Instituto de Matemática e Estatística.

1. Bioinformática - Teses. 2. Histocompatibilidade - Teses. I. Silva, Dilson. II. Sena, Alexandre da Costa. III. Universidade do Estado do Rio de Janeiro. Instituto de Matemática e Estatística. IV Título.

CDU 573:004

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta dissertação, desde que citada a fonte

_____ _____
Assinatura Data

Bruna Garcia Cruz Ribeiro

Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo Laboratório de HLA da UERJ

Dissertação apresentada como requisito parcial para obtenção do título de Mestre ao Programa de Pós-Graduação em Ciências Computacionais, da Universidade do Estado do Rio de Janeiro.

Aprovada em 17 de Abril de 2017.

Banca Examinadora:

Prof. Dr. Dilson Silva (Orientador)
Instituto de Matemática e Estatística - UERJ

Prof. Dr. Alexandre da Costa Sena (Coorientador)
Instituto de Matemática e Estatística - UERJ

Prof. Dr. Igor Machado Coelho
Instituto de Matemática e Estatística - UERJ

Prof. Dr. Luis Cristóvão de Moraes Sobrino Pôrto
Policlínica Piquet Carneiro - UERJ

Prof. Dr. Sotiris Missailidis
Instituto de Tecnologia em Imunobiológicos - FIOCRUZ

Rio de Janeiro

2017

AGRADECIMENTO

Ao meu marido Guilherme, pelo apoio em todos os momentos e pela compreensão nos dias em que tive que me dedicar de maneira exclusiva a este trabalho.

Aos meus pais, que sempre me mostraram que o conhecimento ajuda a construir uma vida.

Aos meus orientadores, Prof. Dr. Dilson Silva e Prof. Dr. Alexandre da Costa Sena, pela compreensão e o apoio dados.

Ao Prof. Dr. Leandro Augusto Justen Marzulo, por ter dado suporte para a conclusão deste trabalho.

À equipe do Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro, em especial à Juliana Fernandes Cardoso, que se mostrou disposta e interessada no meu trabalho.

Aos membros da banca examinadora, pela disponibilidade e contribuições.

RESUMO

RIBEIRO, Bruna Garcia Cruz. *Estudo e proposta de sistema para o processo de tipificação HLA praticado pelo Laboratório de HLA da UERJ*. 2017. 128 f. Dissertação (Mestrado em Ciências Computacionais) - Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2017.

A bioinformática, ao propor o estudo da aplicação de técnicas da informática na área da biologia, tem proporcionado enormes avanços no campo da biologia molecular. Toda essa evolução tem gerado um grande volume de dados biológicos que necessitam cada vez mais da utilização da computação para análise e armazenamento dessa informação. Nesse contexto, o Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro (HLA-UERJ), com a colaboração do Laboratório de Computação de Alto Desempenho (LabCAD), também da UERJ, tem buscado empregar a tecnologia em suas atividades com o objetivo de aumentar a qualidade dos serviços prestados. Desta forma, faz-se necessário que os fluxos de trabalho do HLA-UERJ sejam mapeados a fim de possibilitar o melhor entendimento dos processos atuais. O presente trabalho, com a utilização de técnicas de fluxograma, apresenta o mapeamento das atividades de tipificação HLA exercidas pelo HLA-UERJ e oferece sugestões de melhoria para o controle e gestão dos procedimentos atuais, apresentando uma proposta de implementação de um sistema de controle de *workflow* em uma arquitetura de alta disponibilidade, com os objetivos principais de dirimir riscos operacionais e ampliar a capacidade de análise e rastreabilidade dos dados gerados.

Palavras-chave: Mapeamento. Bioinformática. HLA-UERJ. LabCAD. Histocompatibilidade.

ABSTRACT

RIBEIRO, Bruna Garcia Cruz. *Study and proposal of a system for the HLA typing process practiced by the HLA Laboratory of UERJ*. 2017. 128 f. Dissertação (Mestrado em Ciências Computacionais) - Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, 2017.

Bioinformatics, by proposing the study of computer science techniques for application in the field of biology, has provided enormous advances in the molecular biology field. All this evolution has generated a large volume of biological data which needs increasingly use computer for analysis and storage of this information. In this context, the Histocompatibility and Cryopreservation Laboratory of Rio de Janeiro State University (HLA-UERJ), with the collaboration of the High Performance Computing Laboratory (LabCAD) also UERJ, has sought to use the technology in their activities with the aim to improve the quality of services provided. Thus, it is necessary to sistematize its workflows to allow a better understanding of the current processes. This study, using flowchart techniques, presents the mapping of the HLA typing activity praticed by HLA-UERJ and offers suggestions for improvement in the control and management of current procedures, presenting a workflow management system implementation proposal in a high availability architecture, with the main objectives to settle operational risks and increase the analysis capacity and traceability of the data generated.

Keywords: Mapping. Bioinformatics. HLA-UERJ. LabCAD. Histocompatibility.

LISTA DE FIGURAS

Figura 1 - Desenvolvimento do banco de dados público *GenBank* em conexão com alguns marcos da bioinformática 15
Figura 2 - Crescimento do número de bases e sequências registradas no *GenBank* de 1982 até o momento presente 31
Figura 3 - Representação do fluxo de informação em sistemas biológicos 38
Figura 4 - Representação dos mapas genômicos 40
Figura 5 - Alinhamento de quatro sequências de nucleotídeos envolvendo 55 caracteres. (a) grupo de sequências não alinhadas, cada sequência ocupando uma linha individual. (b) grupo de sequências alinhadas, com os caracteres idênticos dispostos em uma mesma coluna 43
Figura 6 - Diferenças entre alinhamento local e global. (a) duas sequências de nucleotídeos de tamanhos diversos amostradas e alinhadas por algoritmos diferentes. (b) no alinhamento local, a região altamente similar é encontrada e o resto é excluído do resultado final. (c) no alinhamento global, as duas sequências são alinhadas por completo 45
Figura 7 - O MHC humano no braço curto do cromossomo 6 49
Figura 8 - Evolução do número de doadores cadastrados no REDOME por ano até janeiro de 2017 52
Figura 9 - Fontes e frequências de erro no processamento do espécime diagnóstico .. 56
Figura 10 - Total de computadores em uso no Brasil 59
Figura 11 - Dispositivos móveis *wireless* conectáveis à *internet* em uso no Brasil 60
Figura 12 - Exemplo de microplaca e seu mapa de identificação. (a) Exemplo de microplaca com 96 poços. (b) Exemplo de um mapa produzido e utilizado pelo laboratório para identificação das amostras na microplaca de PCR-SSO 68

Figura 13 - Exemplo de mapa de localização e identificação das amostras na microplaca. (a) Exemplo de um mapa produzido e utilizado pelo laboratório para identificação das amostras na microplaca antes de ser submetida ao termociclador. (b) Exemplo de um mapa de identificação das amostras na microplaca refeito após as amostras serem submetidas ao termociclador e algumas serem descartadas 70
Figura 14 - Fluxo do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler* 74
Figura 15 - Fluxo do subprocesso de média resolução do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler* 75
Figura 16 - Fluxo do subprocesso de alta resolução do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler* 75
Figura 17 - Representação esquemática do padrão de arquitetura MVC adotado pelo *Ruby on Rails* 79
Figura 18 - Representação gráfica simplificada do sistema proposto 84
Figura 19 - Diagrama esquemático dos modelos propostos 85
Figura 20 - Tabela do modelo `User` 88
Figura 21 - Tabela do modelo `Tipo` 90
Figura 22 - Tabela do modelo `Fase` 91
Figura 23 - Tabela do modelo `Analise` 93
Figura 24 - Tabela do modelo `AnaliseFase` 95
Figura 25 - Proposta de tela para alteração do fluxo das análises 96
Figura 26 - Tabela do modelo `Solicitante` 97
Figura 27 - Tabela do modelo `Amostra` 98
Figura 28 - Tabela do modelo `AnaliseAmostra` 100
Figura 29 - Proposta de tela para a criação do mapa de identificação e localização das amostras 101
Figura 30 - Esquema gráfico da arquitetura de alta disponibilidade proposta 103
Figura 31 - Tela de Login do protótipo 115
Figura 32 - Tela Inicial do protótipo para o usuário comum 116
Figura 33 - Tela Inicial do protótipo para o usuário administrador 116
Figura 34 - Tela do protótipo para trocar a própria senha 117

Figura 35 - Tela do protótipo para listar usuários 117
Figura 36 - Tela do protótipo para criar novo usuário 118
Figura 37 - Tela do protótipo para exibir usuário 119
Figura 38 - Tela do protótipo para editar usuário 119
Figura 39 - Tela do protótipo para trocar a senha do usuário 120
Figura 40 - Tela do protótipo para listar tipos de análise 120
Figura 41 - Tela do protótipo para criar tipos de análise 121
Figura 42 - Tela do protótipo para exibir tipos de análise 121
Figura 43 - Tela do protótipo para editar tipos de análise 122
Figura 44 - Tela do protótipo para criar fase 122
Figura 45 - Tela do protótipo para exibir fase 123
Figura 46 - Tela do protótipo para editar fase 123
Figura 47 - Tela do protótipo para listar solicitantes 124
Figura 48 - Tela do protótipo para criar solicitantes 124
Figura 49 - Tela do protótipo para exibir solicitantes 125
Figura 50 - Tela do protótipo para editar solicitantes 125
Figura 51 - Tela do protótipo para listar análises 126
Figura 52 - Tela do protótipo para criar análise 126
Figura 53 - Tela do protótipo para exibir análise 127
Figura 54 - Tela do protótipo para editar análise 128

LISTA DE TABELAS

Tabela 1 - Vinte aminoácidos mais comumente encontrados em proteínas 36
Tabela 2 - Resumo das cinco principais famílias de plataformas de sequenciamento NGS 48
Tabela 3 - Especificação técnica do computador de controle do sistema HiSeq 2500 (Illumina®) 72
Tabela 4 - Especificação técnica de um computador do HLA-UERJ 73
Tabela 5 - Principais tipos de dados suportados pelo Rails 80
Tabela 6 - Validações mais comuns fornecidas pelo Rails 81
Tabela 7 - Associações fornecidas pelo Rails 82
Tabela 8 - Ações padrão de um controle criadas automaticamente na geração de um *scaffold* 83
Tabela 9 - Descrição dos campos do modelo `User` 89
Tabela 10- Descrição dos campos do modelo `Tipo` 90
Tabela 11- Descrição dos campos do modelo `Fase` 91
Tabela 12- Descrição dos campos do modelo `Analise` 93
Tabela 13- Descrição dos campos do modelo `AnaliseFase` 95
Tabela 14- Descrição dos campos do modelo `Solicitante` 97
Tabela 15- Descrição dos campos do modelo `Amostra` 98
Tabela 16- Descrição dos campos do modelo `AnaliseAmostra` 100

LISTA DE SIGLAS

ABH	Associação Brasileira de Histocompatibilidade
ARPANET	*Advanced Research Projects Agency Network*
ASP	*Active Server Pages*
BISTIC	*Biomedical Information Science and Technology Initiative Consortium*
CAPES	Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
cDNA	*Complementary Deoxyribonucleic Acid*
CIB	*Center for Information Biology*
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CPU	*Central Processing Units*
DDBJ	*DNA Data Bank of Japan*
DNA	*Deoxyribonucleic Acid*
DNS	*Domain Name System*
EBI	*European Bioinformatics Institute*
EMBL	*European Molecular Biology Laboratory*
EST	*Expressed Sequence Tag*
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FGV	Fundação Getúlio Vargas
FORTRAN	*Formula Translation*
GCG	*Genetics Computer Group*
GPU	*Graphical Processing Units*
HLA	*Human Leukocyte Antigen*
HLA-UERJ	Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro
HUPE	Hospital Universitário Pedro Ernesto
IBM	*International Business Machines*
IFCC	*The International Federation of Clinical Chemistry*
LabCAD	Laboratório de Computação de Alto Desempenho
LABINFO	Laboratório de Bioinformática

LNCC	Laboratório Nacional de Computação Científica
MCT	Ministério da Ciência e Tecnologia
MVC	*Model-View-Controller*
NCBI	*National Center for Biotechnology Information*
NGS	*Next-Generation Sequencing*
NIH	*National Institutes of Health*
PCR	*Polymerase Chain Reaction*
PDB	*Protein Data Bank*
PIR	*Protein Information Resource*
REDOME	Registro Nacional de Doadores de Medula Óssea
SIB	*Swiss Institute of Bioinformatics*
SINAPAD	Sistema Nacional de Processamento de Alto Desempenho
SGBD	Sistema de Gerenciamento de Banco de Dados
SGW	Sistema de Gerenciamento de *Workflow*
SRS	*Sequence Retrieval System*
SSO	*Specific Sequence of Oligonucleotides*
TI	Tecnologia da Informação
UERJ	Universidade do Estado do Rio de Janeiro
UGCDFA	Unidade de Genômica Computacional Darcy Fontoura de Almeida
UNICAMP	Universidade de Campinas
URL	*Uniform Resource Locator*
USB	*Universal Serial Bus*
WASPaLM	*The World Association of Societies of Pathology and Laboratory Medicine*
WWW	*World Wide Web*

SUMÁRIO

INTRODUÇÃO 14

1 BIOINFORMÁTICA 21

1.1 Definições 21

1.2 Breve histórico 22

1.3 Bioinformática no Brasil 27

1.4 Bancos de dados biológicos públicos 28

1.5 Desafios da bioinformática 32

2 CONCEITOS DE BIOLOGIA MOLECULAR 36

2.1 Conceitos básicos 36

2.1.1 Proteínas 36

2.1.2 Ácidos Nucleicos 37

2.1.3 Fluxo da informação em sistemas biológicos 38

2.2 Sequenciamento de DNA 39

2.3 Alinhamento de sequências 43

2.4 Sequenciamento de nova geração (NGS) 46

2.5 Sistema HLA 48

2.5.1 Transplante de Medula Óssea 51

3 MEDICINA LABORATORIAL 53

3.1 Laboratórios diagnósticos 53

3.2 Vantagens da tecnologia da informação 56

3.3 A importância do mapeamento de processos 60

4 PROCESSO DE TIPIFICAÇÃO HLA DO HLA-UERJ 63

4.1 Caracterização do ambiente de estudo 63

4.1.1 Laboratório HLA-UERJ 63

4.1.2 Atividades do Laboratório HLA-UERJ 64

4.1.3 Exames de Tipificação HLA do HLA-UERJ 65

4.2 Detalhamento e mapeamento do processo atual 67

4.3 Identificação de oportunidades de melhoria 76

4.3.1 Ruby on Rails 77
4.3.2 Proposta para um Sistema de Controle de *workflow* em Ruby on Rails 84
4.3.3 Proposta de arquitetura de alta disponibilidade 102

CONCLUSÕES E TRABALHOS FUTUROS 106

REFERÊNCIAS 108

APÊNDICE - Telas do Protótipo 115

INTRODUÇÃO

Desde que a estrutura do DNA foi desvendada em 1953, o campo da biologia molecular tem testemunhado enormes avanços. Com o aumento da capacidade de se manipular sequências de DNA, uma grande quantidade de dados foi e tem sido gerada em diferentes laboratórios espalhados ao redor do mundo [1].

Os anos de 1997 e 1998 podem ser considerados marcos na área, pois nesses dois anos foram publicados o sequenciamento dos genomas de dois organismos vivos relevantes na biologia, o do *Caenorhabditis elegans* e o do *S. cerevisiae*. No ano de 2000, outros dois organismos, o *Arabidopsis thaliana* e o *Drosophila Melanogaster* tiveram seu sequenciamento concluído. No ano seguinte, foi publicado o primeiro esboço do genoma humano e em seguida outros genomas tão importantes quanto foram sendo sequenciados e publicados, como o do rato *Mus musculus* (2002), o do agente causador da malária humana *Plasmodium falciparum* (2002) e de seu vetor o mosquito *Anopheles gambiae* (2002). Em 2004, o do rato castanho *Rattus norvegicus* e em 2005 o do chimpanzé *Pan troglodytes* [2].

A Figura 1 apresenta o crescimento de um dos maiores banco de dados biológicos públicos, o *GenBank*, relacionado a esses principais projetos, que nos últimos 10 anos teve sua base de dados triplicada [3].

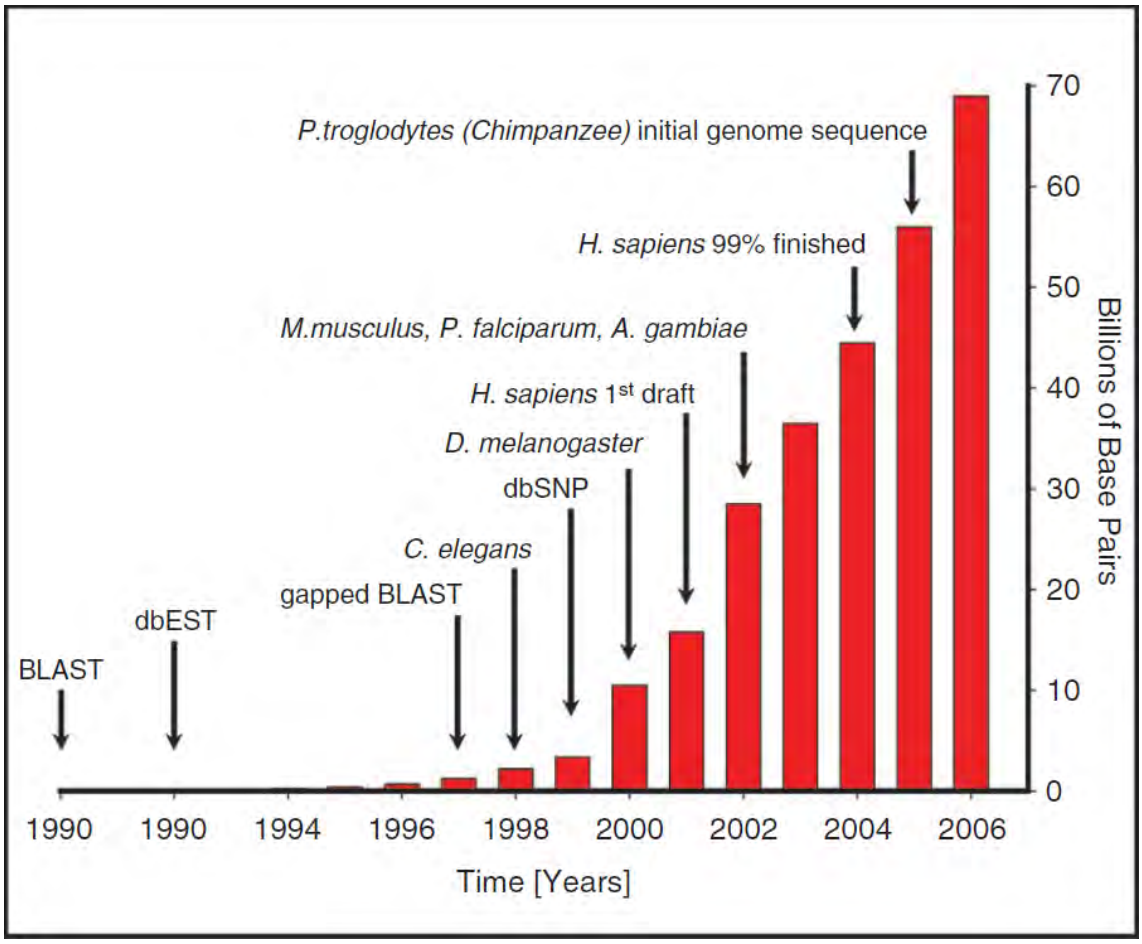
Fica claro que o investimento em projetos de pesquisa ligados a sequenciamento de DNA e análise genômica proporcionou o crescimento do número de bases e sequencias cadastradas em bancos de dados biológicos, abrindo caminho para diferentes possibilidades de análises. Por outro lado, surgiram novos níveis de dificuldade na compreensão, integração e manipulação dos dados, já que estes podem ter formatos diferentes, estarem duplicados ou serem de origem duvidosa.

Essa explosão de informação estimulou a busca repentina por ferramentas computacionais eficientes para gerenciar e analisar os dados gerados. Entretanto, o desenvolvimento dessas ferramentas computacionais dependeu do conhecimento gerado por uma ampla gama de disciplinas como matemática, estatística, informática, tecnologia da informação e biologia molecular [4].

A fusão dessas disciplinas com a biologia resultou em um novo campo de estudo, conhecido como bioinformática. Essa recente área de pesquisa interdisciplinar, que envolve

o uso de tecnologia para armazenamento, recuperação, manipulação e distribuição de informações relacionadas à biologia das macromoléculas, tais como DNA, RNA e proteínas, tem se expandido rapidamente nos últimos anos [4].

Figura 1 - Desenvolvimento do banco de dados público *GenBank* em conexão com alguns marcos da bioinformática



Fonte: [2], f. 27.

De forma geral, o maior objetivo da bioinformática é proporcionar melhor entendimento a respeito de uma célula viva e seu funcionamento em nível molecular, já que, através da análise das sequências e estruturas moleculares, novas idéias e perspectivas podem surgir. A razão pela qual isso ocorre é que o fluxo de informação genética é ditado pelo dogma central da biologia, no qual o DNA é transcrito para o RNA, que por sua vez é traduzido para proteínas. As proteínas são as principais responsáveis pelas funções celulares e é através da análise de suas sequências que suas capacidades são determinadas.

Portanto, a resolução de problemas funcionais usando sequências e abordagens, às vezes, estruturais tem se mostrado um esforço produtivo [4].

Torna-se evidente que a utilização de computadores na análise de dados biológicos tem afetado, em todos os aspectos, as ciências biológicas. Pesquisas que costumavam ter início em laboratórios, hoje, já começam no computador, à medida que os cientistas vasculham bancos de dados a procura de novas hipóteses [5].

Nesse contexto, o Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro (HLA-UERJ), com a colaboração do Laboratório de Computação de Alto Desempenho (LabCAD), também da UERJ, tem buscado empregar a tecnologia da informação em suas atividades com o objetivo de aumentar a qualidade dos serviços prestados.

O HLA-UERJ, localizado na Policlínica Piquet Carneiro, atua nas áreas de ensino, pesquisa e assistência, realizando exames de imunogenética pertinentes aos transplantes renais e atuando como colaborador no cadastramento de doadores voluntários do Registro Nacional de Doadores de Medula Óssea (REDOME) [6].

Devido a investimentos e campanhas de sensibilização da população promovidas pelo Ministério da Saúde e órgãos vinculados, como o INCA, o número de doadores voluntários aumentou de forma expressiva nos últimos anos, passando de 12 mil inscritos no ano de 2000 para mais de 4 milhões nos dias atuais. Com isso, o percentual de transplantes de medula realizados por doadores brasileiros localizados no REDOME subiu de 10%, em 2000, para 70% hoje [6]. A partir desse progresso, "o Brasil tornou-se o terceiro maior banco de dados do gênero no mundo, ficando atrás apenas dos registros dos Estados Unidos e da Alemanha [6].

O laboratório também vem sendo um importante aliado na pesquisa referente a detecção e prevenção de rejeições crônicas renais, acompanhando a evolução de pacientes no pós-transplante e influenciando na medicação empregada, visando a terapia que melhor se adéqua ao paciente, primando pela minimização dos custos [6].

Para a realização dos exames, o laboratório aplica técnicas de sequenciamento de DNA e identificação de genótipos do complexo principal de histocompatibilidade humano, denominado HLA [6]. Esses procedimentos tornam-se complexos à medida que demandam a rastreabilidade e a confiabilidade das amostras e o arquivamento dos dados gerados. Atualmente, grande parte dos processos tem sido realizada manualmente.

O uso de tecnologia traz vantagens como maior celeridade em identificar amostras, segurança dos dados, redução de erros, rastreabilidade completa, entre outros. Porém, obstáculos na comunicação entre biólogos e "informatas" devido à falta de conhecimento tecnológico prejudicam a implantação de novos sistemas. Com a finalidade de dirimir essas dificuldades e para que sejam aplicadas técnicas de informatização no fluxo de trabalho existente, é necessário que seja realizado o mapeamento e a documentação dos processos envolvidos nas atividades do laboratório.

Um processo dentro de uma organização nada mais é do que um "grupo de tarefas interligadas logicamente, que fazem uso dos recursos da organização, para gerar resultados definidos, em apoio aos objetivos da organização"[7].

A estrutura de um processo envolve entradas (*inputs*), saídas (*outputs*), tempo, espaço, ordenação, objetivos e valores. Todos esses fatores correlacionados resultam no fornecimento de produtos ou serviços ao cliente, sendo a sua compreensão um fator de extrema importância para o sucesso de qualquer empresa [8].

Fica patente que "é preciso ter um profundo conhecimento das atividades que constituem tantos os processos essenciais da organização como os processos que os apóiam, considerando finalidades, pontos de início e fim, entradas, saídas e influências limitadoras". Logo, faz-se necessário mapear a organização, revisando processos e sugerindo mudanças [9].

Além disso, "mapear um processo é fundamental para verificar como funcionam, todos os componentes de um sistema, facilitar a análise de sua eficácia e a localização de deficiências. É importante, também, o entendimento de qualquer alteração que se proponha nos sistemas existentes e ainda fornecer subsídio ao tomador de decisão para avaliar as atividades que agregam e que não agregam valor à operação"[9].

Em termos de registro e documentação histórica, o mapeamento dos processos também ganha grande importância para que a qualidade dos serviços seja mantida mesmo com a rotatividade de funcionários, ficando o conhecimento armazenado na organização.

Com o mapeamento das atividades de tipificação HLA exercidas pelo HLA-UERJ é possível observar o estado atual do processo e com isso, identificar oportunidades de melhorias. Essas oportunidades de melhorias visam agilizar o serviço prestado pelo laboratório, com a finalidade de incrementar a qualidade e a segurança dos dados e ampliar o escopo de amostras coletadas e analisadas.

OBJETIVOS

O objetivo geral deste trabalho é realizar o mapeamento das atividades de tipificação HLA exercidas pelo HLA-UERJ, oferecendo sugestões de melhoria para o controle e gestão dos procedimentos atuais através de uma proposta de implementação de um sistema de controle de *workflow* em uma arquitetura de alta disponibilidade, para que os riscos operacionais sejam minimizados e a capacidade de análise e rastreabilidade dos dados gerados seja ampliada.

Os seguintes objetivos específicos foram estabelecidos:

- Compreender conceitos de biologia molecular e bioinformática;
- Detalhar o processo de tipificação HLA do HLA-UERJ atual;
- Apresentar o fluxo do processo utilizando uma ferramenta gráfica;
- Identificar oportunidades de melhoria do processo de tipificação HLA do HLA-UERJ atual;
- Apresentar os recursos necessários à implementação de melhorias;
- Promover a integração entre os laboratórios HLA-UERJ e LabCAD;

METODOLOGIA

Para que os objetivos dessa dissertação fossem atingidos, a metodologia escolhida foi a pesquisa participante, por se tratar de "um tipo de pesquisa baseado numa metodologia de observação participante na qual os pesquisadores estabelecem relações comunicativas com pessoas ou grupos da situação investigada com o intuito de serem melhor aceitos"[10].

A pesquisa participante estimula o envolvimento do pesquisador com o objeto de pesquisa e "permite a reafirmação de fatos, facilitada pela vivência de situações específicas. O pesquisador sente intensamente as dificuldades e as facilidades das situações vivenciadas, inclusive as adaptações necessárias para compreender o campo de interrogativas, fruto de novas inquietações que emergem à medida que passa por essas vivências"[11].

Para o levantamento de dados, foram utilizadas, além da observação, técnicas de análise bibliográfica e entrevista aberta. A combinação de métodos pode ser uma boa estratégia para reduzir o risco de um único método escolhido ser inapropriado, já que as limitações de um método podem ser supridas pela associação de outro método [12].

Com o intuito de contextualizar o estudo e introduzir conceitos acerca da biologia molecular e de sua relação com a computação, na primeira fase da pesquisa, foi realizada uma análise bibliográfica tomando como base livros e artigos científicos.

Em um segundo momento, técnicas de observação participante e entrevistas abertas foram utilizadas para que o fluxo de trabalho atual do processo de tipificação HLA do laboratório HLA-UERJ fosse melhor compreendido. Para tal, foram realizadas algumas reuniões nas dependências do laboratório com os técnicos e pesquisadores envolvidos no processo.

A entrevista aberta tem o objetivo de construir uma base de informações pertinentes para um objeto de pesquisa, sendo uma categoria onde o entrevistado é convidado a falar livremente sobre um tema, o entrevistador apenas realiza perguntas para buscar dar mais profundidade às reflexões. A entrevista, quando analisada, precisa incorporar o contexto de sua produção e, sempre que possível, ser acompanhada e complementada por informações provenientes de observação participante [13]

Uma vez identificados os principais pontos e características das atividades de tipificação HLA, foi realizado o mapeamento do processo atual utilizando técnicas de fluxograma. Além disso, o processo foi analisado de forma a identificar oportunidades de melhorias, assim como os recursos necessários à implementação dessas melhorias.

ORGANIZAÇÃO DO TRABALHO

O presente trabalho está organizado conforme segue:

A Introdução traz uma apresentação do trabalho e está dividida nas seguintes seções: Objetivos, Metodologia e Organização do trabalho.

O Capítulo 1, intitulado "Bioinformática", aborda o referencial teórico sobre a origem da bioinformática, sua aplicação e desafios. Está dividido nas seguintes seções: 1.1 Definições, 1.2 Breve histórico, 1.3 Bioinformática no Brasil, 1.4 Bancos de dados biológicos públicos e 1.5 Desafios da bioinformática.

No Capítulo 2, intitulado "Conceitos de biologia molecular", para melhor entendimento do estudo, são introduzidos alguns conceitos de biologia molecular com ênfase em sequenciamento de DNA. Está dividido nas seguintes seções: 2.1 Conceitos básicos, 2.2 Sequenciamento de DNA, 2.3 Alinhamento de sequências, 2.4 Sequenciamento de nova geração (NGS) e 2.5 Sistema HLA.

O Capítulo 3, intitulado "Medicina laboratorial", aborda a questão dos laboratórios diagnósticos, assim como as vantagens da informatização e a importância de mapear processos. Está dividido nas seguintes seções: 3.1 Laboratórios diagnósticos, 3.2 Vantagens da tecnologia da informação e 3.3 A importância do mapeamento de processos.

O Capítulo 4, intitulado "Processo de tipificação HLA do HLA-UERJ", traz uma apresentação do laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro (HLA-UERJ), detalha o processo de tipificação HLA do HLA-UERJ atual e aponta oportunidades de melhoria e os recursos necessários à implementação. Está dividido nas seguintes seções: 4.1 Caracterização do ambiente de estudo, 4.2 Detalhamento e mapeamento do processo atual e 4.3 Identificação de oportunidades de melhoria.

A seção intitulada "Conclusões e trabalhos futuros" apresenta as conclusões obtidas no desenvolvimento deste trabalho, assim como algumas sugestões para a realização de trabalhos futuros.

1 BIOINFORMÁTICA

Este capítulo aborda o referencial teórico sobre a origem da bioinformática, sua aplicação e desafios. Está dividido nas seguintes seções: 1.1 Definições, 1.2 Breve histórico, 1.3 Bioinformática no Brasil, 1.4 Bancos de dados biológicos públicos e 1.5 Desafios da bioinformática.

1.1 Definições

O conceito de bioinformática entrelaça-se e confunde-se um pouco com o conceito de biologia computacional. Atualmente, ambas são disciplinas de grande importância para a resolução de diversos problemas biológicos.

Os Institutos Nacionais de Saúde dos Estados Unidos (NIH) concordam que biologia computacional e bioinformática são áreas de estudo que se originaram de disciplinas tais como matemática, física, ciência da computação, engenharia e biologia. O comitê de definição BISTIC do NIH, em dezessete de julho do ano de 2000, formalizou as definições desses dois termos da seguinte forma [14]:

Bioinformática: pesquisa, desenvolvimento ou aplicação de ferramentas computacionais e abordagens para expandir o uso dos dados biológicos, médicos, comportamentais ou de saúde, incluindo aquelas para adquirir, armazenar, organizar, arquivar, analisar ou visualizar tais dados.

Biologia computacional: desenvolvimento e aplicação de dados analíticos e métodos teóricos, modelagem matemática e técnicas de simulação computacional para o estudo dos sistemas biológicos, comportamentais e sociais.

Existem algumas variações de interpretação dos conceitos de bioinformática e biologia computacional entre os autores.

Segundo Zomaya [15] biologia computacional é o uso de técnicas computacionais para modelar sistemas biológicos de vários níveis de complexidade – atômico, metabólico, celular e patológico e a bioinformática é a área que proporciona a criação de aplicações escaláveis a fim de dirimir essa complexidade.

Já para Gibas e Jambeck [5], a bioinformática é um subconjunto da biologia computacional que aplica a tecnologia da informação na gestão dos dados biológicos a fim de compreender a biologia em si enquanto biologia computacional é a aplicação de técnicas

analíticas quantitativas na modelagem de sistemas biológicos.

Xiong [4] afirma que a bioinformática se limita à análise sequencial, estrutural e funcional de genes e genomas e seus produtos correspondentes enquanto a biologia computacional envolve todas as áreas da biologia que utilizam alguma forma de computação, como exemplo, modelagem matemática de ecossistemas, o que não necessariamente envolve moléculas biológicas.

Apesar de diferirem em suas definições, bioinformática e biologia computacional são campos de estudo interdisciplinares que requerem habilidades de várias outras áreas para tornar possível a coleta, manuseio, armazenamento, análise, interpretação e disseminação de informação biológica [15].

De acordo com Verli [16], a bioinformática pode ser dividida em duas grandes vertentes: a bioinformática tradicional, relacionada a sequências de nucleotídeos e aminoácidos e a bioinformática estrutural, relacionada a química computacional ou modelagem molecular de um ponto de vista mais tridimensional.

Por outro lado, para Xiong [4], o campo da bioinformática pode ser dividido em outros dois subcampos, que são complementares entre si: o do desenvolvimento de ferramentas computacionais e bancos de dados e o da aplicação dessas ferramentas e bancos de dados para gerar conhecimento biológico.

O subcampo do desenvolvimento de ferramentas computacionais incluiria a criação de *software* para a análise de sequências, assim como a criação e manutenção de bancos de dados biológicos. A aplicação dessas ferramentas e bancos de dados se daria nas áreas de análise de sequências, estruturas e funcionamento do DNA [4].

1.2 Breve histórico

A história da bioinformática está diretamente conectada à história da biologia e à história da informática. Muitos avanços na área da biologia molecular foram realizados antes mesmo do termo bioinformática existir.

A redescoberta dos trabalhos de Gregor Mendel sobre os fatores da hereditariedade, coincidentemente, de forma simultânea e independente, por três pesquisadores de países diferentes, Hugo de Vries (Holanda), Carl Correns (Alemanha) e Erich Von Tschermak (Áustria), proporcionou o desenvolvimento da genética como ciência. As leis de Mendel passaram despercebidas na literatura científica durante 35 anos, só sendo reconhecidas

em 1900 [17].

Nas duas décadas seguintes, os avanços no campo da genética consistiram em entender os mecanismos da hereditariedade a nível molecular. Podemos destacar a introdução dos cromossomos como portadores dos genes e a descoberta da ligação gênica [17].

Na década de 1930, Arne Tiselius aprimorou a eletroforese em gel, que consiste em uma técnica de separação de partículas de acordo com a sua carga elétrica e peso molecular utilizando uma matriz de gel submetida a um campo elétrico. Esse trabalho posteriormente em 1948, rendeu a Tiselius um Prêmio Nobel de Química [18].

O aperfeiçoamento das técnicas de análise molecular do material genético trouxe várias conquistas importantes para a genética na década de 1940. Em 1944, Oswald Avery, Colin MacLeod e Maclyn McCarty apresentaram o DNA como o mensageiro da informação genética e não as proteínas, como se pensou durante um tempo, o que veio a ser confirmado pelos experimentos de Alfred Hershey e Martha Chase em 1952 [18].

A confirmação de que o DNA seria o material da hereditariedade acabou por intensificar os estudos acerca de sua estrutura e em 1953, os pesquisadores Francis Crick e James Watson, baseados em imagens do DNA conseguidas através da técnica de difração de raios X por Rosalind Franklin e Maurice Wilkins, apresentaram um modelo em dupla hélice para a estrutura da molécula de DNA. Crick, Watson e Wilkins receberam o Prêmio Nobel em 1962 em reconhecimento aos seus trabalhos. Franklin já havia falecido em 1958 [18].

Avanços nos processos de análise laboratorial possibilitaram em 1955, a determinação da primeira sequência completa de uma proteína, a insulina, por Frederick Sanger, que ganhou o Prêmio Nobel de Química em 1958 por essa descoberta. Embora os estudos de bioquímica das proteínas necessitassem de instrumentos cada vez mais precisos, os computadores não eram vistos como colaboradores decisivos nesse processo. Eram usados apenas para lidar com os tipos de dados numéricos que eram gerados em grande quantidade pelos pesquisadores [19].

O lançamento da primeira linguagem em alto nível, FORTRAN, pela IBM, em 1957, impulsionou a área da biologia computacional, pois era uma linguagem relativamente fácil de aprender e que não carecia de conhecer maiores detalhes da arquitetura dos computadores para se escrever um programa, sendo bem recebida pelos cientistas [19].

No início da década de 1960, os computadores ficaram mais disponíveis para os

pesquisadores acadêmicos. Segundo pesquisas realizadas na época, 15% das faculdades e universidades dos Estados Unidos possuíam pelo menos um computador no campus e as principais universidades voltadas para pesquisas estavam adquirindo os chamados computadores de segunda geração, que eram baseados em transistores, para substituir os antigos, que utilizavam tubos de vácuo [19].

Na mesma época, com o apoio dos NIH, *National Science Foundation, National Aeronautics and Space Administration* e IBM uma série de programas em FORTRAN foram escritos por Margaret Dayhoff na tentativa de determinar sequências de aminoácidos de proteínas [19].

Em 1965, Dayhoff desenvolveu o primeiro banco de dados de sequências de proteínas, o *Atlas of Protein Sequence and Structure*, reconhecido como o primeiro esforço para sistematizar o conhecimento acerca da informação genética. Por ter sido uma das pioneiras no uso de computadores no estudo da biologia molecular, é considerada a fundadora da bioinformática [16].

As conquistas científicas também se davam na área da ciência da computação. Em 1969, foi criada a ARPANET, primeira rede de computadores que permitia a troca de dados entre universidades e alguns centros de pesquisa dos Estados Unidos. Inicialmente conectava quatro nós, progredindo para 15 nós em 1971 e para 35 nós em 1973. Em 1977, a rede já possuía 111 computadores e em 1983, já possuía 4000 computadores [20].

Quanto mais sequências de DNA e proteínas tornavam-se disponíveis, mais o interesse em desenvolver programas de computadores para analisar essas sequências de todas as formas aumentava. Em 1970, o primeiro algoritmo para comparação de sequências de aminoácidos foi construído por Needleman e Wunsh. Em ciência da computação, essa abordagem ficou conhecida como programação dinâmica [21].

Embora os computadores não estivessem desenvolvidos a ponto de permitir a representação gráfica da estrutura de uma proteína, pelo menos não com uma velocidade aceitável, em 1971, foi criado o PDB no *Brookhaven National Laboratory*, um banco de dados para armazenar as coordenadas da estrutura em um arquivo de computador específico. O PDB possuía menos de 400 entradas até o final da década de 1980 [5].

No ano de 1972, Paul Berg conseguiu sintetizar a primeira molécula de DNA recombinante derivada de dois fragmentos diferentes de DNA. Esse foi o ponto de partida para os processos de clonagem. No ano seguinte, em 1973, a primeira molécula de DNA

clonada foi obtida a partir de genes de rã por Stanley Cohen e Herbert Boyer [22].

Em 1977, foi publicada por Sanger a primeira sequência completa de DNA do genoma de um bacteriófago, o *phi-x174*, com aproximadamente 5.375 nucleotídeos [23]. Por esse feito, Sanger recebeu o seu segundo Prêmio Nobel de Química.

Em 1978, o termo bioinformática foi utilizado pela primeira vez por Paulien Hogeweg ao estudar os processos informáticos em sistemas biológicos [2].

O desenvolvimento da bioinformática se deu de forma lenta nos anos seguintes e somente em 1980 foi fundada, na Califórnia, a *IntelliGenetics*, primeira empresa voltada para o desenvolvimento de software cientifico que lançou nesse mesmo ano, a primeira coleção de softwares voltados para a análise de sequências de DNA e proteínas, o *IntelliGenetics Suíte* [2].

Em 1981, acompanhando a necessidade de se comparar apenas uma parte das sequências biológicas ao invés das sequências completas, Smith e Waterman desenvolveram um algoritmo para alinhamentos locais, que buscava apenas as regiões da sequência com maior similaridade. Esse algoritmo atualmente é o principal algoritmo utilizado nos programas voltados para alinhamento local [16].

No mesmo ano, a IBM introduziu no mercado o primeiro computador pessoal de baixo custo, o IBM 5150. Isso possibilitou a inclusão do computador em centenas de laboratórios que não estavam familiarizados com essa tecnologia, o que permitiu maior agilidade na análise dos dados brutos. Conforme o processamento de dados ia se tornando mais rápido e barato, os biologistas moleculares faziam um uso cada vez maior desse recurso [24].

Em 1982, a Universidade de Wisconsin criou o *Genetics Computer Group* (GCG), cujo principal produto era o *Wisconsin Suíte*, um pacote de aplicativos similar ao *IntelliGenetics Suíte*, também voltado para biologia molecular. Em um primeiro momento, tanto o *IntelliGenetics Suíte* quanto o *Wisconsin Suíte* eram compostos por pequenos programas controlados via linha de comando. Mais tarde, foi desenvolvida uma interface gráfica para o *Wisconsin Suíte*, o que facilitou a sua operação pelos usuários. Com o passar do tempo, o *IntelliGenetics Suíte* veio a desaparecer do mercado e o *Wisconsin Suíte* mudou seu nome para *Accelrys GCG* [2].

Em 1983, Kary Mullis desenvolveu a técnica de reação de polimerização em cadeia (PCR) que possibilitou aos pesquisadores a produção de bilhões de cópias de uma

seqüência de DNA específica de forma automatizada. O primeiro sequenciador de DNA automático foi desenvolvido por Leroy Hood e seus colaboradores e lançado em 1986 pela empresa americana *Applied Biosystems*. Atualmente, a PCR é uma técnica indispensável e muito comumente usada nos laboratórios para estudos genéticos, no diagnóstico de doenças hereditárias e infecciosas, para testes de paternidade e na ciência forense investigativa. Mullis recebeu o Prêmio Nobel de Química em 1993 por esse feito [25].

No ano de 1986, Thomas Roderick cunhou o termo "genômica" para descrever a disciplina científica que estuda o genoma [2]. Ainda em 1986, foi fundado o *SWISS-PROT*, um banco de dados que provê não só a sequência, mas também a descrição da função da proteína, sua estrutura de domínio e outras informações. A partir de 1987 passou a ser mantido, colaborativamente, pelo Departamento de Bioquímica Médica da Universidade de Genebra e pela Biblioteca de dados do EMBL, atualmente conhecida como Instituto Europeu de Bioinformática (EMBL-EBI) [26].

Em 1988, foi fundado, nos Estados Unidos, o Centro Nacional para Informação Biotecnológica, o NCBI, com o objetivo de criar sistemas automatizados que servissem de fonte de informação para as áreas de biologia molecular, bioquímica e genética [27]. No mesmo ano, W. Pearson e David Lipman desenvolveram o FASTA, um algoritmo capaz de realizar buscas mais rápidas por similaridade entre sequências em um banco de dados [21].

Outro algoritmo, o BLAST, para comparação de sequências, tão rápido quanto o FASTA, foi desenvolvido por Altschul, Gish, Miller, Myers e Lipman em 1990. Atualmente, o BLAST é provavelmente o algoritmo mais utilizado para análise de sequências no mundo [21]. Ainda em 1990, foi lançado oficialmente o Projeto Genoma Humano, um esforço internacional para mapear todos os genes que formam o corpo humano.

Em 1991, foi desenvolvida, por Craig Venter, a técnica de etiqueta de sequência expressa (EST), a qual consiste em separar apenas uma parte do clone de uma sequência complementar de DNA (cDNA) que seja suficiente para que a proteína seja identificada. As ESTs se tornaram amplamente aceitas pela facilidade e rapidez de obtenção a um custo reduzido. Como reação ao enorme crescimento das ESTs geradas, foi criado o banco de dados dbEST no NCBI para armazenar todas as ESTs publicamente conhecidas [2].

Em 1995, foi publicado o sequenciamento completo do primeiro genoma bacteriano, o da *Haemophilus Influenzae*, dando início a uma nova era na pesquisa genômica [2].

Nos anos seguintes, o progresso no campo da bioinformática contribuiu para que fossem finalizados sequenciamentos completos de genomas de diversas espécies de organismo. O sequenciamento completo referente ao genoma humano foi publicado em 2003, finalizando-se assim, o Projeto Genoma Humano [21].

1.3 Bioinformática no Brasil

A bioinformática foi introduzida no Brasil em 1999, por João Meidanis e João Setúbal, cientistas do Laboratório de Bioinformática da Universidade de Campinas (UNICAMP) responsáveis pelo desenvolvimento de ferramentas tecnológicas para o sequenciamento genético com base na internet, no âmbito do Programa Genoma-FAPESP, que resultou no sequenciamento completo do DNA da bactéria *Xylella fastidiosa* [28]. Em 2000, a ciência brasileira ganhou destaque internacional, quando a pesquisa sobre a bactéria *Xylella* saiu na capa da *Nature*, uma das mais importantes revistas científicas do mundo. Era a primeira vez que a revista publicava um trabalho realizado no Brasil em sua capa.

Antes mesmo do início do Programa Genoma-FAPESP, Meidanis e Setubal lançaram em 1994 um livro sobre bioinformática, o "Introdução a Biologia Computacional", com a finalidade de disseminar a matéria e com isso, conquistar novos alunos para a área. Segundo eles, o livro lançado foi o terceiro sobre bioinformática em todo o mundo [28].

O pioneirismo da dupla estimulou o investimento na área e diversos centros de bioinformática foram sendo criados pelo país. No ano de 2000, com o apoio do Programa de Biotecnologia e Recursos Genéticos do Ministério de Ciência e Tecnologia (MCT), foi dado início ao Laboratório de Bioinformática (LABINFO), no Laboratório Nacional de Computação Científica (LNCC), com o objetivo de ser um laboratório de referência em nível nacional, baseado em pesquisa interdisciplinar, envolvendo biologia, matemática e computação.

Em 2001, a CAPES considerou a área da bioinformática uma das áreas que careciam de apoio, já que, por possuir caráter multidisciplinar, ela sofria com a tendência dos alunos em buscar apenas as disciplinas mais tradicionais [29]. No ano seguinte, a própria CAPES divulgou apoio à área através da criação de novos cursos de doutorado na Universidade de São Paulo (USP) e na Universidade de Minas Gerais (UFMG) com suporte financeiro por cinco anos [30]. Nos anos seguintes, outros cursos foram surgindo

em diversas universidades e faculdades em vários estados brasileiros.

A primeira empresa de bioinformática do Brasil, a *Scylla Bioinformática*, foi criada em 2002 por Meidanis com o objetivo de diminuir a distância entre a pesquisa acadêmica e a sua aplicação no mercado e dois anos depois, em 2004, foi fundada a Associação Brasileira de Bioinformática e Biologia Computacional (AB3C), uma sociedade científica dedicada ao avanço do entendimento de seres vivos através de métodos formais, multidisciplinares e quantitativos de larga escala. Desde sua criação, a AB3C promove anualmente o *X-Meeting*, uma conferência brasileira com alcance internacional para difundir conhecimento na área da bioinformática. A AB3C também apoia vários cursos na área de ensino e aplicação de técnicas em Bioinformática e Biologia Computacional.

Ainda no início dos anos 2000, o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) investiu três milhões de reais no aperfeiçoamento e implementação de núcleos de bioinformática em todo o país. Esse investimento ajudou a ampliar a infraestrutura de pesquisa dos programas de sequenciamento e analise de genoma que eram patrocinados pelo MCT [31].

Alguns anos depois, em 2008, podemos destacar a inauguração da Unidade de Genômica Computacional Darcy Fontoura de Almeida (UGCDFA), um centro de excelência e de referência nacional em sequenciamento de nova geração destinada a atender aos projetos genomas de todo o país [32].

Em 2015, foi instalado no LNCC o maior computador da América Latina, com enorme capacidade de processamento de informações numéricas. O supercomputador, batizado com o nome de Santos Dumont, ou SDumont, estará integrado ao Sistema Nacional de Processamento de Alto Desempenho (SINAPAD) através da rede acadêmica brasileira, a RNP-Rede Nacional de Pesquisa [33]. Será de grande valia para as pesquisas em genômica que estão atualmente em curso no Brasil.

1.4 Bancos de dados biológicos públicos

O desenvolvimento das técnicas de biologia molecular resultou em um aumento significativo da geração de dados genômicos e proteômicos, tais como sequências e estruturas moleculares, mapas genômicos físicos e genéticos, polimorfismos, informação bibliográfica, propriedades químicas moleculares, entre outras [34].

Mediante ao volume de informação crescente que se fazia presente nos laboratórios,

surgiu a necessidade da estruturação e organização desses dados biológicos, a fim de agilizar o seu processamento e promover maior intercambio das informações entre os centros de pesquisas espalhados pelo mundo. A princípio, os dados foram sendo armazenados em bibliotecas locais nos próprios centros de pesquisa. Posteriormente, com o avanço da tecnologia e a facilidade do acesso ao meio eletrônico com o advento da *Internet*, essas bibliotecas foram migradas para o meio digital e transformadas, em sua grande parte, em arquivos públicos para consulta. Hoje, existem inúmeros bancos de dados voltados para a pesquisa biológica que podem ser acessadas de qualquer lugar através da *Internet* [34].

Bancos de dados nada mais são do que arquivos de computador usados para armazenar e organizar informação de maneira que seja possível a sua recuperação rápida e fácil utilizando algum critério de busca. Originalmente, eram todos do tipo de arquivo simples, os chamados *flat files*, que são basicamente arquivos de texto com vários registros separados por algum delimitador textual. Essa característica os tornava pouco gerenciáveis e demandava uso intenso de memória do computador, já que para realizar consultas era necessário que o computador percorresse o arquivo inteiro em busca do dado [4].

Para facilitar o acesso e recuperação dos dados e melhorar a utilização dos recursos computacionais, foram desenvolvidos programas conhecidos como Sistemas de Gerenciamento de Banco de Dados (SGBD). Esses sistemas não só contém os registros brutos, mas também instruções operacionais que ajudam na identificação das conexões entre esses registros, o que torna a busca do dado mais ágil. São dois os tipos: os relacionais e os orientados a objeto[4].

Os SGBDs ditos relacionais utilizam uma série de tabelas relacionadas entre si para estruturar os dados, onde as colunas representam a identificação do dado e as linhas os dados em si. Os SGBDs ditos orientados a objeto são os voltados para a abordagem de dados complexos e são baseados em relacionamento hierárquico dos dados, os tratando como objetos [4].

Apesar de os bancos de dados relacionais e orientados a objeto serem mais eficientes do que os baseados em arquivos simples, já que são mais fáceis de gerenciar e proporcionam buscas mais rápidas, eles não são bem aceitos no campo dos bancos de dados biológicos. Um dos motivos seria a vantagem da possibilidade de se manipular os arquivos de texto sem a necessidade de ferramentas complexas, o que reduz o custo e amplia o intercâmbio de arquivos entre os cientistas. Já existem sistemas capazes de indexar os registros de um

banco de dados baseado em arquivo simples para acelerar as consultas [2].

Segundo Xiong [4], os bancos de dados biológicos podem ser divididos em três categorias: primários, secundários e especializados. Bancos de dados primários são os que contêm dados originais, em sua forma bruta, sem qualquer tipo de processamento, que são submetidos pela comunidade científica. São exemplos dessa categoria o *GenBank* e o PDB. Bancos de dados secundários são os que contêm dados já processados, tanto computacionalmente quanto manualmente, baseados na informação contida nos bancos primários. São exemplos dessa categoria o *SWISS-PROT* e o PIR. Bancos de dados especializados são os que contem informação muito específica a respeito de alguma matéria em particular. Um exemplo dessa categoria é o banco de dados *Flybase*, que é utilizado para a análise do genoma de *Drosophilas*.

Anualmente, o jornal *Nucleic Acids Research*, publica uma lista atualizada, com as respectivas URLs, de todos os bancos de dados biológicos disponíveis no mundo. A consulta à lista também pode ser realizada através do endereço eletrônico: http://www.oxfordjournals.org/our_journals/nar/database/c/.

Os três maiores bancos biológicos primários públicos de sequências de ácidos nucleicos são os *GenBank*, EMBL e DDBJ, os quais estão disponíveis na *Internet*. Apesar de terem formatos diferentes, existe uma rotina de sincronização entre os três bancos, ou seja, todos contem as mesmas sequências [4].

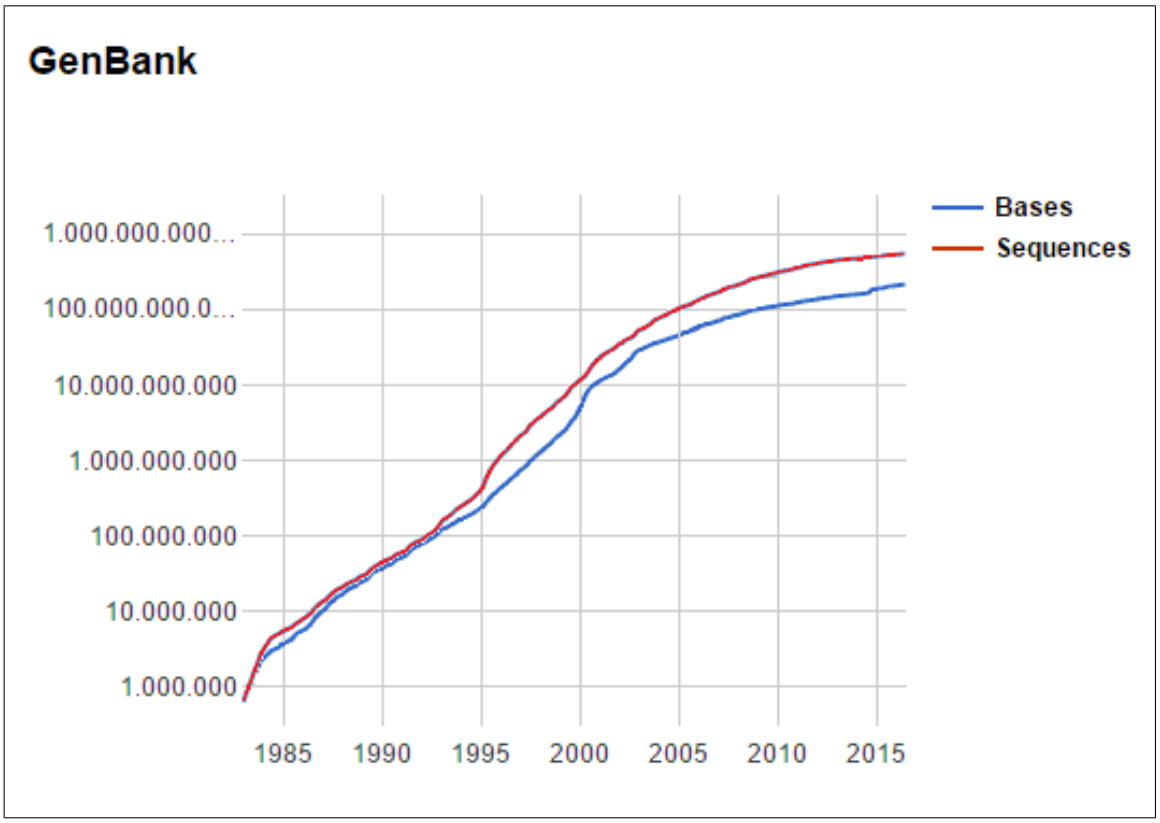
O *GenBank* é disponibilizado pelo NCBI, nos Estados Unidos e foi criado principalmente a partir de dados de sequências enviados por laboratórios individuais e pelo envio em massa de dados de projetos de sequenciamento de larga escala. A cada dois meses são liberadas novas sequências, que podem ser obtidas através do endereço eletrônico: ftp://ftp.ncbi.nih.gov [2].

O registro das sequências, no *GenBank*, pode ser realizado por qualquer pessoa via página da WWW, utilizando a plataforma *Bankit* ou, quando a sequência for extensa, por *email*, utilizando a ferramenta *Sequin*, ambas as interfaces são fornecidas e mantidas pelo NCBI [2].

A consulta aos registros do *GenBank* pode ser realizada através do sistema *Entrez*, um portal *web* mantido pelo NCBI. Esse portal não só permite pesquisas simples e complexas às bases de ácido nucleicos, mas a todas as outras bases mantidas também pelo NCBI [2]. A Figura 2 apresenta o crescimento do número de bases e de sequências

registradas no *GenBank* desde 1982 até o presente momento.

Figura 2 - Crescimento do número de bases e sequências registradas no *GenBank* de 1982 até o momento presente



Fonte: [3].

O EMBL está localizado no EBI e é o maior da Europa, e o DDBJ está no Centro de Informação de Biotecnologia (CIB), no Japão e é o maior da Ásia. Embora o formato do DDBJ seja o mesmo utilizado pelo *GenBank*, o do EMBL se diferencia um pouco. A principal diferença é no uso de um código de duas letras no início de cada linha, ao invés de palavras inteiras, além de outras pequenas modificações na organização dos campos dos dados [2].

A consulta aos bancos DDBJ e EMBL é realizada através do Sistema de Recuperação de Sequências (SRS), que foi desenvolvido pelo EBI para gerenciar bancos de dados biológicos primários e secundários. O EMBL utiliza o portal EBI-SRS e o DDBJ o portal DDBJ-SRS [2].

Para a publicação científica de novas sequências, é pré-requisito que essa nova

sequência esteja registrada no *GenBank* ou em qualquer um dos bancos associados a ele, tais como o EMBL e o DDBJ [2].

O principal e único banco de dados que centraliza as estruturas tridimensionais das macromoléculas biológicas é o PDB. Esse banco de dados utiliza o formato de arquivo simples para representar e arquivar o nome da proteína, autores, detalhes experimentais, estrutura secundaria, cofatores e coordenadas atômicas das macromoléculas, que são determinadas pelas técnicas de cristalografia de raios X e Ressonância Magnética Nuclear. A própria interface *web* do banco de dados PDB fornece ferramentas de consulta e manipulação de imagens [4].

Os principais bancos biológicos secundários públicos de sequências de proteínas são o *SWISS-PROT*, o *TrEMBL* e o PIR [34].

O PIR é um banco de dados não redundante, anotado e totalmente classificado. Foi criado a partir do *Atlas of Protein Sequence and Structure* para suportar pesquisas nas áreas genômica e proteômica [34].

O *SWISS-PROT* é um dos mais importantes bancos de sequências de proteínas e está localizado no Instituto Suíço de Bioinformática (SIB). Por ser curado manualmente, isto é, cada registro é verificado por um especialista, possui alta qualidade e mínima redundância [2].

Como os especialistas do SIB não conseguem acompanhar o ritmo de crescimento do *SWISS-PROT*, foi criado o banco de dados *TrEMBL* com o objetivo de suplementá-lo ao receber automaticamente os registros de novas entradas. Sendo assim, o *TrEMBL* contém todas as sequências de ácidos nucleicos que ainda não foram analisadas pelos especialistas nem incluídas no *SWISS-PROT*, portanto a qualidade do *TrEMBL* é inferior [2].

Recentemente os bancos *SWISS-PROT*, *TrEMBL* e PIR foram unidos em um único banco, chamado *UniProt*, com o objetivo de desenvolver um catalogo central de informação de proteínas. O *UniProt*, apesar de ter enorme abrangência, consegue manter anotações de alta qualidade [4].

1.5 Desafios da bioinformática

A união entre a informatização e a moderna biologia molecular contribuiu para o surgimento dessa nova área do conhecimento, a bioinformática. Vários projetos já foram

finalizados ou estão em fase de conclusão ao redor do mundo, sendo os responsáveis por uma enorme quantidade de informação biológica gerada [35].

Segundo Verli [16], o crescimento deste volume de informação está longe de cessar. Vários estudos em andamento ainda têm muito a agregar ao nosso conhecimento do funcionamento de sistemas biológicos, o que potencializa aplicações terapêuticas e biotecnológicas. Porém, tudo isso exigirá cada vez mais avanços em *hardware, software*, em modelagem e análise de dados.

Computacionalmente, trabalhar com sequências é menos custoso do que trabalhar com estruturas de proteínas. É possível o uso de genomas inteiros em processos que envolvem comparações entre sequências, identificação de padrões em sequências ou construção e anotações de genomas, construção de redes de sistemas biológicos e caracterização de relações evolutivas. A manipulação de estruturas demanda maior poder de processamento, portanto só é possível o uso de uma ou apenas um punhado de estruturas de forma simultânea para, por exemplo, obter modelos 3D ou identificar o modo de interação de moléculas [16].

Por ser uma área que está sempre em desenvolvimento, a bioinformática exige constante atualização tecnológica, pois programas tendem a se tornar rapidamente ineficientes ou mesmo obsoletos. A fim de suprir essa necessidade, algumas áreas tem tido maior concentração de esforços da comunidade científica, tais como o uso de CPUs (unidades de processamento central) e GPUs (unidades de processamento gráfico) para processamento de dados [16].

As CPUs, que são partes dos computadores responsáveis pela execução das instruções estabelecidas pelos programas, têm se tornado cada vez mais complexas, confiáveis e baratas, com isso, é possível a construção de modelos mais precisos de fenômenos biológicos cada vez mais complexos. Hoje, a maioria das CPUs utilizadas já possui múltiplos núcleos. Essa característica permite que os programas sejam adaptados para que o cálculo seja distribuído pelos núcleos de processamento, diminuindo o tempo de resposta. Vários programas de bioinformática já são compatíveis com o multi-processamento [16].

As GPUs, que são microprocessadores inicialmente desenvolvidos para processamento gráfico, estão geralmente localizadas nas placas de vídeo de computadores e podem chegar a possuírem centenas ou milhares de núcleos de processamento. Essa característica passou a ser explorada na computação científica, implicando em grande aumento na velo-

cidade de manipulação de dados. Porém, a possibilidade do uso da computação em GPU dependerá das características do problema em questão e da eficiência e portabilidade do código empregado [16].

Mesmo com a tecnologia avançando concomitantemente às descobertas biológicas, ainda há muito mais dados gerados do que informação extraída.

O armazenamento desses dados biológicos em bancos de dados públicos tem se tornado cada vez mais comum, com isso esses bancos de dados têm crescido exponencialmente [5]. Assim, o grande desafio da bioinformática é armazenar e manter o enorme volume de informação com o auxílio de ferramentas computacionais [4].

Contudo, a missão de integrar toda essa informação é árdua, pois existem diversos bancos de dados espalhados pelo mundo, cada um com sua representação, sistemas de gerenciamento e formatos específicos.

A redundância de informação também é um grande problema que afeta os bancos de dados primários, seja pela inclusão repetida de sequências idênticas ou pela falha do sistema de gerenciamento do banco em detectar dados duplicados, o que pode tornar o banco de dados excessivamente grande, prejudicando a performance de consultas [4].

Outra questão que afeta os bancos de dados é a inconsistência na anotação da sequência, onde a mesma sequência pode ser encontrada sobre diferentes nomes, resultando em múltiplas entradas. Pesquisadores têm buscado desenvolver sistemas de nomeação capazes de prover nomes consistentes e sem ambiguidade para todos os genes e proteínas [4].

Como as atividades em bioinformática envolvem conceitos tanto de computação quanto de biologia, a comunicação entre esses profissionais pode ser um tanto quanto difícil. Essa dificuldade demandou o aparecimento de um profissional que compreendesse tanto problemas biológicos quanto sistemas computacionais, o chamado "bioinformata".

Segundo Araujo [35], esses novos profissionais devem ser capacitados para desenvolver programas computacionais que permitam reconhecer sequências de genes, predizer a configuração tridimensional de proteínas, identificar inibidores de enzimas, organizar e relacionar informação biológica, agrupar proteínas homólogas, estabelecer árvores filogenéticas, analisar experimentos de expressão gênica, entre outros.

O "bioinformata" deve ser altamente multidisciplinar, ou seja, deve ser capaz de lidar com diversos profissionais de diferentes áreas, tais como química, física, entre outras.

Sendo assim, diante de todos esses requisitos necessários, fica evidente que este profissional se enquadra em um perfil específico e bem peculiar [35].

2 CONCEITOS DE BIOLOGIA MOLECULAR

Neste capítulo, para melhor entendimento do estudo, são introduzidos alguns conceitos de biologia molecular com ênfase em sequenciamento de DNA. Está dividido nas seguintes seções: 2.1 Conceitos básicos, 2.2 Sequenciamento de DNA, 2.3 Alinhamento de sequências, 2.4 Sequenciamento de nova geração (NGS) e 2.5 Sistema HLA.

2.1 Conceitos básicos

A área de pesquisa em biologia molecular é basicamente dedicada ao entendimento da estrutura e funcionamento de proteínas e ácidos nucleicos.

2.1.1 Proteínas

Uma proteína é uma biomolécula cuja cadeia é formada de moléculas mais simples chamadas aminoácidos. Encontramos vinte aminoácidos diferentes, que são os mais comuns em proteínas [1], conforme listados na Tabela 1.

Tabela 1 - Vinte aminoácidos mais comumente encontrados em proteínas

	Nome	**Sigla e Abreviatura**			**Nome**	**Sigla e Abreviatura**	
1	Alanina	A	Ala	11	Metionina	M	Met
2	Cisteína	C	Cys	12	Asparagina	N	Asn
3	Ácido aspártico	D	Asp	13	Prolina	P	Pro
4	Ácido glutâmico	E	Glu	14	Glutamina	Q	Gln
5	Fenilalanina	F	Phe	15	Arginina	R	Arg
6	Glicina	G	Gly	16	Serina	S	Ser
7	Histidina	H	His	17	Treonina	T	Thr
8	Isoleucina	I	Ile	18	Valina	V	Val
9	Lisina	K	Lys	19	Triptofano	W	Trp
10	Leucina	L	Leu	20	Tirosina	Y	Tyr

Fonte: [1].

A forma tridimensional de uma proteína determina a sua função, já que os tipos de moléculas aos quais uma proteína pode se ligar vão depender diretamente de sua forma. Por exemplo, a forma de uma proteína pode ser tal que a possibilite se ligar com várias

cópias idênticas de si mesma, construindo, digamos, uma linha de cabelo. Ou a forma pode ser tal que algumas outras moléculas se liguem à proteína e assim comecem a trocar átomos, nesse caso a proteína funcionaria como um catalisador [1].

As proteínas são produzidas em uma estrutura celular chamada ribossomo. Nos ribossomos, as proteínas são sintetizadas utilizando-se as informações contidas nos ácidos nucleicos [1].

2.1.2 Ácidos Nucleicos

Os organismos vivos contem dois tipos de ácidos nucleicos: os ácidos ribonucleicos (RNA) e os ácidos desoxirribonucléicos (DNA) [1].

Assim como as proteínas, uma molécula de DNA é uma cadeia de moléculas mais simples, chamadas de nucleotídeos. Cada nucleotídeo é composto por um açúcar, um fosfato e uma base, que pode ser de quatro tipos: adenina (A), guanina (G), citosina (C) e timina (T). Em uma célula humana, as moléculas de DNA possuem centenas de milhares de nucleotídeos [1].

O DNA é uma molécula de segmento duplo, amarrados em uma estrutura helicoidal, a famosa dupla hélice descoberta por *James Watson* e *Francis Crick* em 1953. Cada base em um segmento forma uma ligação com uma base no outro segmento, respeitando-se que a base A é sempre ligada à base T e a base C é sempre ligada à G, ou seja, as bases A e T e as base C e G são ditas complementares, sendo conhecidas como pares de base de *Watson-Crick* [1].

Um ponto positivo dessa estrutura é que é possível deduzir a sequência de nucleotídeos de um segmento através do outro utilizando a complementação reversa. Por exemplo, dado o segmento s = AGACGT, para conseguirmos o seu complemento inverso, primeiramente, precisamos inverter s. Assim, obteremos s' = TGCAGA. Após, basta substituir cada base por seu complemento, obtendo $\overline{s}$ = ACGTCT [1].

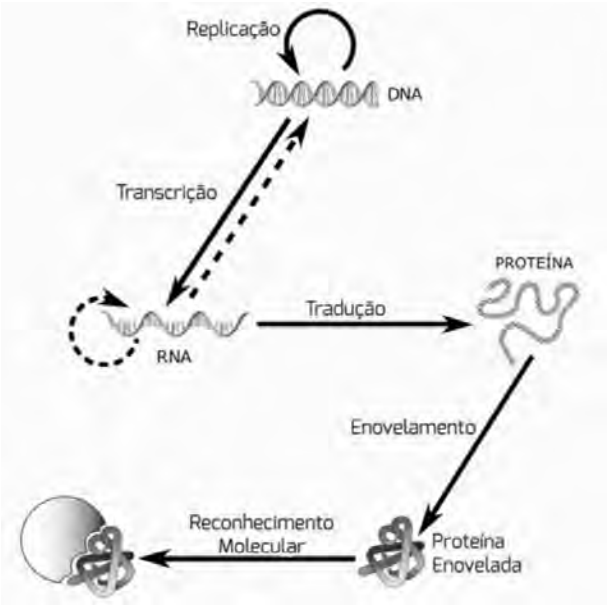
Uma molécula de RNA é bem parecida com a de DNA, com algumas diferenças de composição e de estrutura. Na molécula de RNA, o açúcar encontrado é o ribose, enquanto que a molécula de DNA contém o açúcar desoxirribose. A base timina (T) não é encontrada na RNA, ao invés dela encontramos a uracila (U), que assim como a timina, se liga a adenina (A). Outra característica é que a molécula de RNA não possui estrutura em dupla hélice. Partes de uma molécula de RNA podem ligar-se a outras

partes da mesma molécula por complementaridade, portanto a estrutura tridimensional do RNA pode apresentar uma variedade maior do que a do DNA. Uma outra diferença entre o DNA e o RNA é que, enquanto o DNA executa essencialmente uma função, a de codificar a informação, existem diferentes tipos de RNAs na célula desempenhando diferentes funções [1].

2.1.3 Fluxo da informação em sistemas biológicos

As informações necessárias para construir cada proteína ou RNA encontrado nos organismos está nas moléculas de DNA [1]. A Figura 3 apresenta a representação gráfica do fluxo da informação em sistemas biológicos.

Figura 3 - Representação do fluxo de informação em sistemas biológicos



Fonte: [16], f. 15.

Segundo Verli [16], "... tais dados não são usados diretamente, mas através de uma molécula intermediária, o RNA (mais precisamente o RNAm), sintetizado por um processo denominado transcrição (uma molécula de ácido nucleico é transcrita em outra molécula de ácido nucleico). Esta molécula de RNAm irá servir como molde para a síntese de proteínas, em um processo chamado de tradução (uma molécula de ácido nucleico é traduzida em uma molécula de proteína). As proteínas, assim expressas, irão reger a

maioria dos fenômenos relacionados à função dos organismos e à perpetuação da vida (embora diversos outros processos sejam modulados por outras biomoléculas)".

O esquema tradicional sofreu a adição do processo de enovelamento de reconhecimento molecular devido ao seu caráter fundamental para a manifestação da função gênica [16]. O enovelamento é o processo pelo qual a proteína assume a sua forma.

2.2 Sequenciamento de DNA

O sequenciamento de DNA é um processo realizado com o objetivo de se extrair as informações genéticas contidas nas moléculas de DNA. Essas informações são determinadas de acordo com a sequência dos pares de bases A, T, C e G. Cada trecho contíguo ao longo do DNA que contém a informação necessária para construir uma proteína ou uma molécula de RNA é conhecido como gene [1].

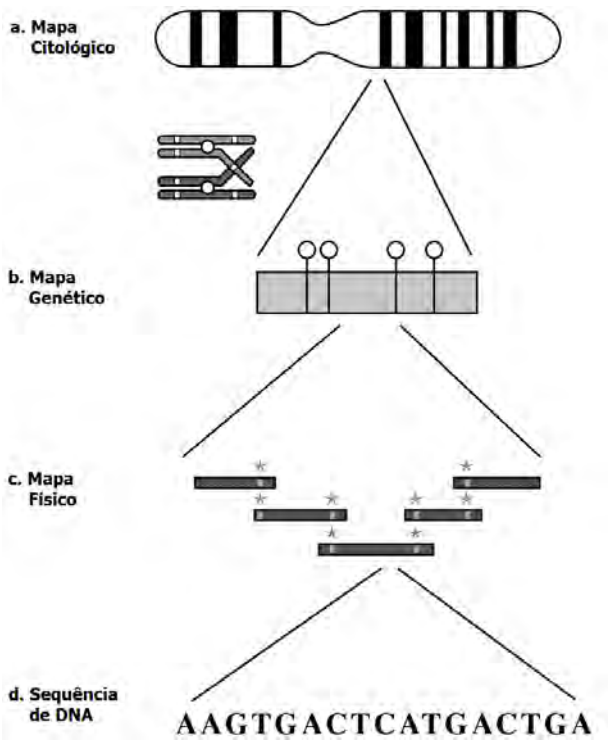
Como a proteína é uma cadeia de aminoácidos, para definir uma proteína é preciso especificar cada aminoácido contido na cadeia. Cada aminoácido é especificado por uma sequência de três bases, chamada de triplete ou códon. Já que existem 64 tripletes nucleotídeos possíveis e apenas 20 aminoácidos a especificar, é possível que diferentes tripletes correspondam ao mesmo aminoácido. Por exemplo, tanto o AAG como o código AAA especificam a lisina. Por outro lado, três dos códons possíveis não especificam nenhum aminoácido e são usados para sinalizar o fim de um gene [1].

Cada célula do organismo possui algumas longas moléculas de DNA e cada molécula dessa é chamada cromossomo. Um conjunto de cromossomos contidos em uma célula é chamado genoma. O número de cromossomos em um genoma varia de espécie para espécie. No ser humano, são encontrados 46 cromossomos em cada célula [1].

O primeiro passo para a compreensão de uma estrutura genômica é através do mapeamento do genoma, que é um processo de identificação de localizações relativas de genes ou mutações em um cromossomo. Uma abordagem de baixa resolução para mapear genomas é descrever a ordem e as distâncias relativas dos marcadores genéticos. Os marcadores genéticos são partes identificáveis de um cromossomo cujos padrões de herança podem ser seguidos. Assim sendo, os mapas genéticos identificam as posições relativas dos marcadores genéticos num cromossoma e baseiam-se na frequência com que os marcadores são herdados em conjunto. Quanto mais próximos estão os dois marcadores genéticos, é mais provável que eles sejam herdados juntos, não sendo separados em um evento de

cruzamento genético. Além dos mapas genéticos, existem também outros tipos de mapas genômicos, como mapas físicos e mapas citológicos, que descrevem genomas em diferentes níveis de resolução [4]. A Figura 4 apresenta os diversos mapas genômicos, representando diferentes níveis de resolução para descrever um genoma utilizando marcadores genéticos.

Figura 4 - Representação dos mapas genômicos



Fonte: [4], f. 244.

Os mapas físicos são mapas de locais de marcos identificáveis em um DNA genômico, independentemente dos padrões de herança. Como a distância é expressa em unidades físicas, eles são mais precisos e confiáveis do que os mapas genéticos. Os mapas citológicos referem-se a padrões de bandas observados em cromossomos corados. As faixas claras e escuras observáveis são os marcadores visualmente distintos em um cromossomo. Um marcador genético pode estar associado a uma banda ou região cromossômica específica. Os padrões de bandas, no entanto, nem sempre são constantes e estão sujeitos a alterações dependendo da extensão da contração cromossômica. Assim, mapas citológicos podem ser considerados de muito baixa resolução e, portanto, um pouco imprecisos [4].

Mapas citológicos são obtidos microscopicamente. Os mapas genéticos (barra cinza) são obtidos através de experimentos de cruzamento genético, no qual cada recom-

binação genética é analisada. Os mapas físicos são obtidos a partir de clones sobrepostos identificados por hibridação dos fragmentos de clone (barras cinza) com sondas comuns (asteriscos cinza) [4].

Como as moléculas de DNA são muito longas, é necessária a aplicação de alguma técnica de corte. Um dos métodos utilizados é o de *shotgun*, no qual cada molécula individual é quebrada em vários locais aleatórios e os fragmentos resultantes são filtrados e selecionados para posterior processamento, em particular para cópia ou clonagem, também chamado de processo de amplificação de DNA. Esse processo é necessário para que seja possível a repetição dos experimentos, já que sempre é possível produzir novo material a partir do que está sendo analisado. Em seguida, é obtida uma coleção de fragmentos clonados que correspondem a pedaços contíguos aleatórios da sequência de DNA purificada [1].

Para a obtenção dos fragmentos clonados, pode-se aplicar tanto a técnica de clonagem de DNA quanto a da Reação em Cadeia da Polimerase (PCR). Na técnica da clonagem de DNA utiliza-se a própria natureza para a obtenção. Um pedaço de DNA é inserido no genoma de um organismo, hospedeiro ou vetor, então deixamos o organismo se multiplicar. Consequentemente, o pedaço de DNA se multiplica junto. O DNA produzido desta forma é denominado recombinante [1].

Uma alternativa à técnica de clonagem de DNA é utilizar a técnica de Reação em Cadeia da Polimerase (PCR). O DNA polimerase é uma enzima que catalisa o alongamento de uma única cadeia de DNA, desde que haja um DNA modelo ao qual esta única cadeia está ligada. Nucleótidos complementares são adicionados aos da cadeia molde até que ambas as cadeias tenham o mesmo tamanho e formem uma cadeia dupla normal de DNA. O pequeno trecho de DNA de cadeia dupla inicialmente necessário para a polimerase começar seu trabalho é chamado de *primer* ou iniciador [1].

A PCR consiste basicamente numa repetição alternada de duas fases: uma fase em que o DNA de cadeia dupla é separado em duas cadeias simples por calor e uma fase em que cada cadeia simples assim obtida, é convertida numa dupla cadeia por adição de um *primer* e ação da polimerase. Cada repetição duplica o número de moléculas. Depois de repetições suficientes, a quantidade de material produzido é suficientemente grande para realizar experimentos adicionais, graças ao crescimento exponencial deste procedimento de duplicação [1].

A leitura dos pares de uma sequência de DNA é feita com uma técnica conhecida como eletroforese em gel, que se baseia na separação de moléculas pelo seu tamanho. O processo envolve um meio gel e um campo elétrico forte. As moléculas de DNA ou RNA são carregadas em solução aquosa e se movem de acordo com a direção definida pela ação do campo elétrico. O meio gel faz com que elas se movam lentamente, com velocidade inversamente proporcional ao seu tamanho. Todas as moléculas são inicialmente colocadas em uma extremidade de um bloco de gel. Depois de algumas horas, as moléculas menores terão migrado para a outra extremidade do bloco, enquanto que as moléculas maiores permanecerão ainda perto do ponto de partida. Por interpolação, os tamanhos relativos das moléculas podem ser calculados de forma bem aproximada [1].

Neste tipo de experimento, as moléculas de DNA podem ser marcadas com isótopos radioativos para que o gel possa ser fotografado, produzindo um registro gráfico das posições no final de uma rodada. Além de utilizado no sequenciamento de DNA, este processo é utilizado na determinação de comprimentos de fragmentos de restrição. Uma alternativa é usar corantes fluorescentes em vez de isótopos radioativos. Um feixe de laser pode rastrear os corantes e enviar as informações diretamente para um computador, evitando o processo fotográfico por completo. As primeiras máquinas de sequenciamento foram construídas usando esta técnica [1].

As bases de DNA ou RNA podem ser lidas utilizando o processo descrito anteriormente com a seguinte técnica. Dada uma molécula de DNA, é possível obter todos os fragmentos dela que terminam em cada posição onde um A aparece. Da mesma forma, todos os fragmentos que terminam em T, em G ou em C podem ser obtidos. Assim, obtemos quatro tubos de ensaio diferentes, um para cada base. Os fragmentos em cada tubo terão comprimentos diferentes. Por exemplo, suponha que a peça de DNA original seja a seguinte: GACTTAGATCAGGAAACT. Os fragmentos que terminam em T são GACT, GACTT, GACTTAGAT, e toda a sequência em si. Assim, se separarmos todos os fragmentos dos tubos por tamanho, saberemos precisamente a composição de base da sequência de DNA original. Alguns erros podem ocorrer na leitura de um filme gel, porque às vezes as marcas são desfocadas, especialmente perto das bordas do filme [1].

2.3 Alinhamento de sequências

O alinhamento de DNA busca uma série de caracteres individuais que se encontram na mesma ordem das sequências de DNA analisadas, utilizando técnicas de comparação entre duas ou mais sequências biológicas. [16].

Durante o alinhamento, as sequências são dispostas em linha com cada caractere biológico correspondendo a uma coluna do alinhamento. A partir dessa organização inicial, os algoritmos específicos buscarão a melhor correspondência entre as sequências, podendo ser criados espaços entre os caracteres para que as sequências terminem com o mesmo comprimento, facilitando a visualização da similaridade [16]. A Figura 5 apresenta um alinhamento de quatro sequências de nucleotídeos envolvendo 55 caracteres, onde os caracteres idênticos são identificados por asteriscos e os nucleotídeos ausentes em determinadas sequências são substituídos por hifens para identificar eventos de inserção/deleção.

Os métodos de alinhamento possibilitam a obtenção de informações a respeito da relação evolutiva entre organismos, indivíduos, genes ou entre sequências diversas. Podem ainda, fornecer informações a respeito da função e da estrutura de sequências biológicas [16].

Figura 5 - Alinhamento de quatro sequências de nucleotídeos envolvendo 55 caracteres. (a) grupo de sequências não alinhadas, cada sequência ocupando uma linha individual. (b) grupo de sequências alinhadas, com os caracteres idênticos dispostos em uma mesma coluna

```
a.                    10        20        30        40        50
                .....|.....|.....|.....|.....|.....
Sequência 01    ATATACTTGATCGATCAGCATCAGCTAGTCGAAGTTTGAGCATGCATGTGTGATG
Sequência 02    CTTGTTCCATCAGCTTCAGCTCGTCGAAGGAGCTGCATGTGTGA
Sequência 03    CTTGTTCCATCAGCTTCAGCTCGTCGAAGGAGCATGCA
Sequência 04    ATATACTTGATCGATTAGCTTCAGCTAGTGGAGCCAGTATGTGTGTG
                 *             *      **   * *

b.                    10        20        30        40        50
                .....|.....|.....|.....|.....|.....
Sequência 01    ATATACTTGATCGATCAGCATCAGCTAGTCGAAGTTTGAGCATGCATGTGTGATG
Sequência 02    -----CTTGTTCCATCAGCTTCAGCTCGTCGAAG---GAGC-TGCATGTGTGA--
Sequência 03    -----CTTGTTCCATCAGCTTCAGCTCGTCGAAG---GAGCATGCA---------
Sequência 04    ATATACTTGATCGATTAGCTTCAGCTAGT----G---GAGCCAGTATGTGTG-TG
                      **** ** ** *** ****** **     *   ****   * *
```

Fonte:[16], f. 40.

As estratégias de alinhamento podem ser tanto locais quanto globais [4].

No alinhamento global, presume-se que duas sequências a serem alinhadas sejam semelhantes ao longo de todo o seu comprimento. Assim, o alinhamento é realizado do início ao fim com o objetivo de se encontrar o melhor alinhamento possível ao longo de todo o comprimento entre as duas sequências. Este método é mais aplicável para o alinhamento de duas sequências rigorosamente relacionadas e de comprimentos aproximados. Para sequências divergentes e de comprimentos variáveis, este método pode não ser capaz de gerar resultados ótimos, já que não é capaz de reconhecer regiões locais altamente semelhantes entre as duas sequências [4].

Por outro lado, o alinhamento local não assume que as duas sequências em questão possuem semelhança ao longo de todo o comprimento. Ele apenas encontra regiões locais com o maior nível de semelhança entre as duas sequências e alinha estas regiões sem considerar o alinhamento do resto da sequência. Esta abordagem pode ser utilizada para alinhar sequências mais divergentes ou de comprimentos variáveis, objetivando encontrar padrões de conservação nas sequências de DNA ou proteínas [4]. A Figura 6 apresenta um exemplo de um alinhamento de duas sequências de nucleotídeos de tamanhos diferentes e aponta as diferenças entre os alinhamentos local e global.

O alinhamento de duas sequências é realizado utilizando os seguintes métodos [21]: análise de matriz de pontos, algoritmos de programação dinâmica e métodos de palavras, como os utilizados pelos programas FASTA e BLAST.

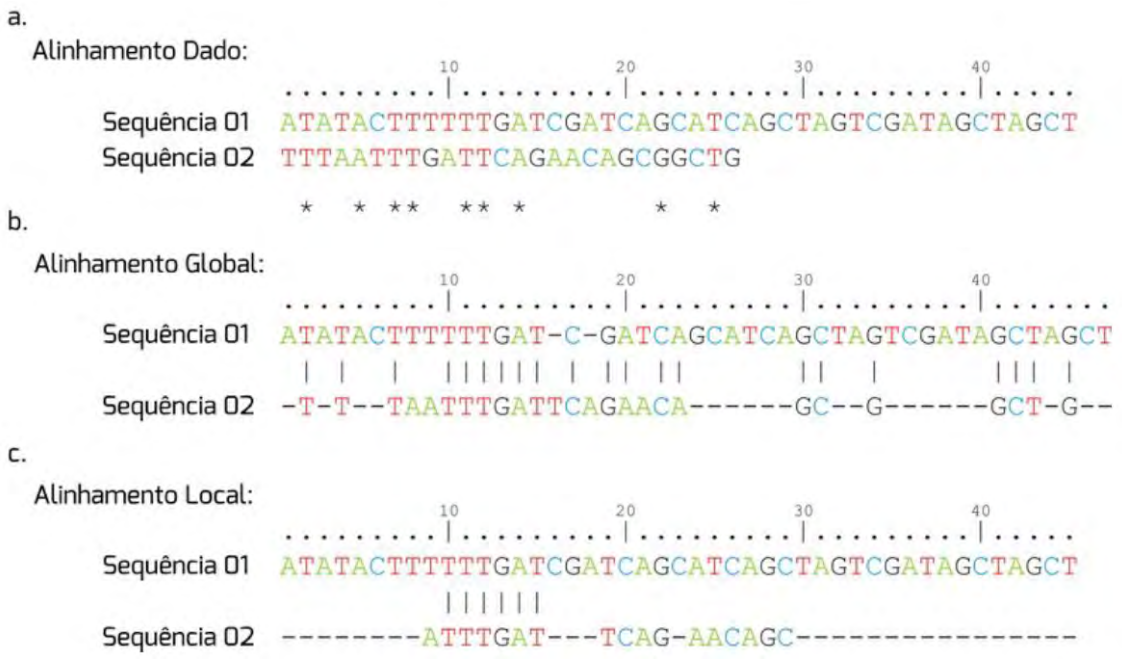
O método de análise de matriz de pontos exibe quaisquer alinhamentos de sequência possíveis como diagonais em uma matriz, sendo assim, a menos que as sequências sejam muito semelhantes, esse método deve ser utilizado primeiro. A análise de matriz de pontos pode prontamente revelar a presença de inserções, deleções e repetições diretas e invertidas, que são mais difíceis de serem apontadas pelos demais métodos. Porém, a limitação principal desse método é a de não mostrar um alinhamento de fato [21].

O método de programação dinâmica gera o alinhamento iniciando nas extremidades das duas sequências, tentando fazer coincidir todos os pares possíveis de caracteres, seguindo um esquema de pontuação para correspondências, desajustes e lacunas. Este procedimento cria uma matriz de números que representa todos os possíveis alinhamentos entre as sequências. O maior conjunto de pontuações sequenciais na matriz define um alinhamento ótimo. Esse método sempre encontra um alinhamento ótimo para um dado

conjunto de variáveis definidas pelo usuário. Entretanto, o método de programação dinâmica pode ser lento devido ao grande número de passos computacionais, que aumentam aproximadamente de acordo com o quadrado ou o cubo dos comprimentos das sequências. O requisito de memória de computador também aumenta conforme o quadrado dos comprimentos das sequências. Assim, torna-se difícil sua utilização para sequências muito longas [21].

Os métodos de palavras, que são utilizados pelos algoritmos FASTA e BLAST, alinham duas sequências muito rapidamente, primeiro procurando por trechos curtos idênticos das sequências (palavras) e em seguida, unindo-as em um alinhamento aplicando o método de programação dinâmica. Esses métodos são rápidos o suficiente para serem adequados a realizarem uma busca em um banco de dados inteiro, objetivando encontrar a sequência que melhor se alinhe com a sequência de entrada [21].

Figura 6 - Diferenças entre alinhamento local e global. (a) duas sequências de nucleotídeos de tamanhos diversos amostradas e alinhadas por algoritmos diferentes. (b) no alinhamento local, a região altamente similar é encontrada e o resto é excluído do resultado final. (c) no alinhamento global, as duas sequências são alinhadas por completo



Fonte: [16], f. 47.

O alinhamento de três ou mais sequências é possível, porém, apesar de conceitualmente simples, é exigida grande capacidade computacional e, dependendo da quantidade de sequências envolvidas, pode ser requerido um longo tempo para processamento [16].

2.4 Sequenciamento de nova geração (NGS)

Os esforços em sequenciar o genoma humano resultou em grandes avanços na automação dos processos e o investimento em tecnologia culminou no surgimento dos métodos chamados *Next-generation sequencing* (NGS), onde os passos de clonagem passaram a ser desnecessários, apesar de também ocorrer a fragmentação aleatória do DNA genômico. Estes novos métodos permitem a obtenção de sequências de forma mais rápida, porém com tamanho menor, variando de algumas dezenas a poucas centenas de pares de base [16].

Estas novas tecnologias constituem várias estratégias que dependem de uma combinação de preparação de modelos, sequenciamento e imagem, e alinhamento do genoma e métodos de montagem. O principal avanço oferecido pela NGS é a capacidade de produzir um volume enorme de dados a baixo custo, em alguns casos, chegando a um bilhão de leituras curtas por execução. Esse recurso expande o domínio da experimentação além de apenas determinar a ordem das bases. Por exemplo, nos estudos de expressão gênica, os *microarrays* estão agora sendo substituídos por métodos de sequenciamentos de RNA (RNA-Seq), que podem identificar e quantificar transcritos raros sem conhecimento prévio do gene em questão e podem fornecer informações sobre *splicing* alternativo e variação de sequência de genes [36].

As plataformas que mais se destacam dentre as tecnologias de nova geração comercialmente disponíveis hoje são: a 454 (Roche), a Illumina (Solexa), a Applied Biosystems SOLiD, a Ion Torrent e a Pacific Biosciences (PacBio) [37].

A 454 Life Sciences (uma empresa Roche) introduziu a tecnologia NGS no mercado. A família de plataformas 454 tem sido usada desde então para uma grande variedade de aplicações e sua capacidade em executar longas leituras se fez especialmente atraente para estudos de microbiomas, uma vez que leituras mais longas geralmente podem ser identificadas com maior acurácia e precisão. A abordagem global da 454 é baseada em pirosequenciamento utilizando amplificação por PCR em emulsão, no qual adaptadores específicos em cada extremidade da sequência são criados utilizando *primers* de PCR e

fixados a esferas minúsculas suspensas em óleo. Uma desvantagem da abordagem 454 é que frequentemente o comprimento dos homopolímeros (trechos de nucleotídeos nos quais todas as bases são idênticas) são identificados erroneamente. Além disso, esta tecnologia costuma ser considerada ineficaz quando comparada com outras tecnologias de sequenciamento NGS, pois é possível a produção de mais sequências com Illumina, SOLiD ou Ion Torrent. No entanto, para algumas aplicações que requerem comprimentos de leitura mais longos, continua a ser a plataforma mais rentável. Embora a tecnologia tenha sido inovadora quando foi introduzida, após 2016, a 454 não será mais suportada [37].

A Illumina produz a família de plataformas mais utilizada, pois uma quantidade maior de dados pode ser gerada de uma maneira mais econômica. Utiliza a abordagem de terminadores marcados reversíveis, de modo que são removidos após cada ciclo de imagem para dar lugar ao próximo nucleotídeo reversível com terminação marcada. A plataforma com a maior saída em geral é a HiSeq 2500, produzindo 4 bilhões de fragmentos de forma pareada com 125 bases para cada leitura em uma única execução [37].

A Applied Biosystems SOLiD não atingiu o mesmo nível de popularidade que as plataformas 454 e Illumina para pesquisas com microbiomas. O processo SOLiD começa com uma fase de PCR em emulsão semelhante à utilizada por 454, porém utiliza a abordagem de sequenciamento por ligação, ao contrário do sequenciamento por síntese que é utilizado nas demais plataformas. Embora possa gerar um grande número de sequências em uma execução, o comprimento de leitura curto persistente limitou muito a sua utilidade [37].

Até 2010, a 454 tinha esculpido um nicho de fornecer leituras mais longas, enquanto a Illumina e a SOLiD tinham demonstrado a capacidade de fornecer um grande número de sequências de uma só vez. Entretanto, surgiu no mercado a necessidade de instrumentos capazes de atender a um escopo de magnitude reduzida. Então, dentro de um curto período de tempo, o 454 GS Junior, o Illumina MiSeq e o SOLiD FlowChip foram lançados e orientados aos laboratórios menores. O Ion Torrent entrou no mercado em 2010 com a máquina de genoma pessoal (PGM), alegando ser a primeira empresa a realmente trazer sequenciamento NGS viável e acessível para laboratórios menores [37].

O sistema Ion Torrent difere de todos os outros sistemas na maneira em que a incorporação de base é detectada. Quando uma base é adicionada a uma cadeia de DNA

crescente um próton é liberado, o que altera ligeiramente o pH circundante. Os microdetectores sensíveis ao pH estão associados aos poços da placa e registram quando essas mudanças ocorrem, o que permite inferir a sequência de cada poço. Uma característica interessante, no entanto, é que os fragmentos mais longos do que aqueles que podem ser totalmente sequenciados através deste sistema são atualmente removidos através de um passo de seleção de tamanho, tornando impossível a sequência de extremidades de fragmentos mais longos. Como o 454, o Ion Torrent também é suscetível a erros relacionados ao homopolímero. A abordagem do Ion Torrent pode ser bastante eficaz para gerar dados de microbioma [37].

A Pacific Biosciences (PacBio) usa uma abordagem de sequenciamento em tempo real em uma única molécula. O sistema de sequenciamento PacBio não envolve nenhuma etapa de amplificação, o que a diferencia dos outros sistemas de sequenciamento NGS [37].

A Tabela 2 apresenta um resumo das características principais da plataformas citadas anteriormente.

Tabela 2- Resumo das cinco principais famílias de plataformas de sequenciamento NGS

Plataformas	**Amplificação clonal**	**Química**	**Comprimento de leitura (média)**
454	PCR em emulsão	Pirosequenciamento (seq por síntese)	700 bp
Illumina	Amplificação em ponte	Terminador de corante reversível (seq por síntese)	300 bp
SOLiD	PCR em emulsão	Ligação em cadeia de oligonucleotídeo 8-mero (seq por ligação)	75 bp
Ion Torrent	PCR em emulsão	Detecção de próton (seq por síntese)	400 bp
PacBio	N/A (molécula única)	Nucleotídeos fluorescentes fosforilados (seq por síntese)	8,500 bp

Fonte: [37].

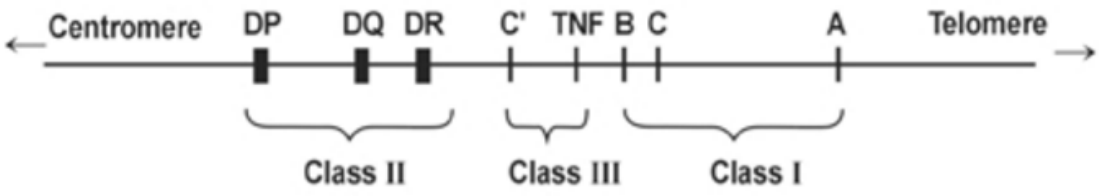
2.5 Sistema HLA

As plataformas NGS inicialmente foram usadas apenas para o sequenciamento genômico, mas também mostraram potencial para fins de pesquisa e diagnóstico [38]. A

análise do genoma de NGS é de suma importância para melhor definição de doenças hematológicas, em termos de patogênese, fatores prognósticos desconhecidos e comportamento de doenças. Além disso, tal sequenciação tem tido grande importância nos domínios da imunologia e da imunoterapia [39].

O principal componente do sistema imunológico é o sistema HLA, também conhecido como antígenos de transplante. Os locus genéticos envolvidos na rejeição de órgãos transplantados são conhecidos como o complexo principal de histocompatibilidade (MHC) e o MHC humano é denominado sistema HLA (Human Leukocyte Antigen). O MHC humano está localizado no braço curto do cromossomo 6 e abrange cerca de 3.600 kilobases de DNA. O MHC humano é dividido em três regiões conforme apresentado na Figura 7 [40].

Figura 7 - O MHC humano no braço curto do cromossomo 6



Fonte: [40], f. 11.

A região de classe I contém os genes HLA-A, HLA-B e HLA-C. A região da classe II contém as sub-regiões HLA-DR, DP e DQ, cada uma delas contendo um gene para uma cadeia alfa e pelo menos um gene para uma cadeia beta. A região de classe III não codifica moléculas de HLA, mas contém os genes que codificam vários componentes do sistema do complemento (C2, C4, B), fatores de necrose tumoral (TNFs) e alguns outros. As moléculas HLA de classe I são expressas na superfície de quase todas as células nucleadas. As moléculas de classe II são expressas apenas em linfócitos B, células apresentadoras de antígenos (monócitos, macrófagos e células dendríticas) e linfócitos T ativados. O papel biológico das moléculas HLA classe I e classe II é apresentar antígenos peptídeos processados [40].

O sistema HLA é conhecido por ser o mais polimórfico em seres humanos. A distribuição e a frequência dos antígenos HLA variam muito entre os diferentes grupos

étnicos. Tem sido postulado que esta diversidade de polimorfismo HLA evoluiu sob pressão seletiva única em diferentes áreas geográficas. Isso pode estar relacionado ao papel das moléculas HLA na apresentação de agentes infecciosos prevalentes nas diferentes áreas do mundo [40].

OS HLA-A, HLA-B e HLA-DR têm sido conhecidos há muito tempo como antígenos de transplante principal. Dados clínicos indicam que a correspondência HLA-C também afeta os resultados clínicos do transplante de células estaminais hematopoéticas, mas o HLA-DQ e o HLA-DP não parecem críticos [40].

Os alelos HLA são designados pelo locus seguido de um asterisco (*), um número de dois dígitos correspondendo à especificidade antigênica e o número de alelos atribuído. Por exemplo, HLA-A*0210 representa o décimo alelo HLA-A2 dentro da família de antígenos HLA-A2 serologicamente definidos [40].

Vários testes clínicos de HLA são realizados em laboratórios a fim de apoiar os programas de transplante, incluindo a tipagem HLA do receptor e a do potencial doador, rastreio e identificação de anticorpos HLA no receptor e detecção de anticorpos no receptor que são reativos com linfócitos de um possível doador, isto é, teste de prova cruzada (*crossmatching*) [40]. A compatibilidade HLA entre doadores e receptores reduz a probabilidade de rejeição e de apresentação da doença do enxerto contra hospedeiro (GVHD) em transplantes [38].

Os benefícios da compatibilidade HLA estão bem estabelecidos no transplante de rim. Existe uma clara relação entre o grau de compatibilidade HLA e a sobrevivência do enxerto renal em transplantes de doadores vivos. Melhores resultados são obtidos de um doador irmão HLA-idêntico do que com doadores pais, irmãos ou filhos HLA-haploidênticos. O transplante de rim de um doador vivo tem uma sobrevivência do enxerto superior ao transplante de um doador falecido (com exceção da combinação de seis antígenos) apesar de um maior grau de incompatibilidade de HLA. Estes resultados favoráveis são provavelmente devido à redução do tempo entre a retirada e o transplante ser reduzido e ao dano renal ser menor [40].

O transplante alogênico de células estaminais hematopoiéticas é utilizado para tratar a malignidade hematológica, anemia aplástica grave, imunodeficiências congênitas graves e doenças metabólicas hereditárias. A medula óssea é a fonte das células estaminais hematopoéticas, porém o uso de sangue de cordão umbilical tem se tornado cada vez

mais popular. A principal vantagem dos transplantes de sangue do cordão umbilical é que a incompatibilidade do HLA é melhor tolerada, tendo menos resultados adversos em transplantes [40].

2.5.1 Transplante de Medula Óssea

O transplante de medula óssea é um tipo de tratamento proposto para algumas doenças que afetam as células do sangue, como as leucemias e os linfomas. Consiste na substituição de uma medula óssea doente, ou deficitária, por células normais da medula óssea, com o objetivo de reconstituição de uma nova medula saudável. É na medula óssea que se localizam as células-tronco hematopoéticas, responsáveis pela geração de todo o sangue (glóbulos vermelhos, glóbulos brancos e plaquetas). Essas são as células substituídas no transplante de medula [41].

As células-tronco hematopoéticas também circulam no sangue periférico (caso estimuladas com medicamento de fator de crescimento), podendo ser coletadas por aférese e no sangue do cordão umbilical, quando são coletadas após o nascimento do bebê. Por isso, o termo "transplante de medula óssea" tem sido substituído por "transplante de células-tronco hematopoéticas" para estes procedimentos [41].

REDOME é o Registro Nacional de Doadores Voluntários de Medula Óssea, um cadastro que reúne informações (nome, endereço, resultados de exames, características genéticas, etc) de voluntários à doação de medula para pacientes que precisam do transplante. Um sistema informatizado cruza as informações genéticas dos doadores voluntários cadastrados no REDOME com as dos pacientes que precisam do transplante. Quando é verificada a compatibilidade, a pessoa é convocada para efetivar a doação [42]. É possível se cadastrar como doador voluntário de medula óssea nos hemocentros localizados em todos os estados do país.

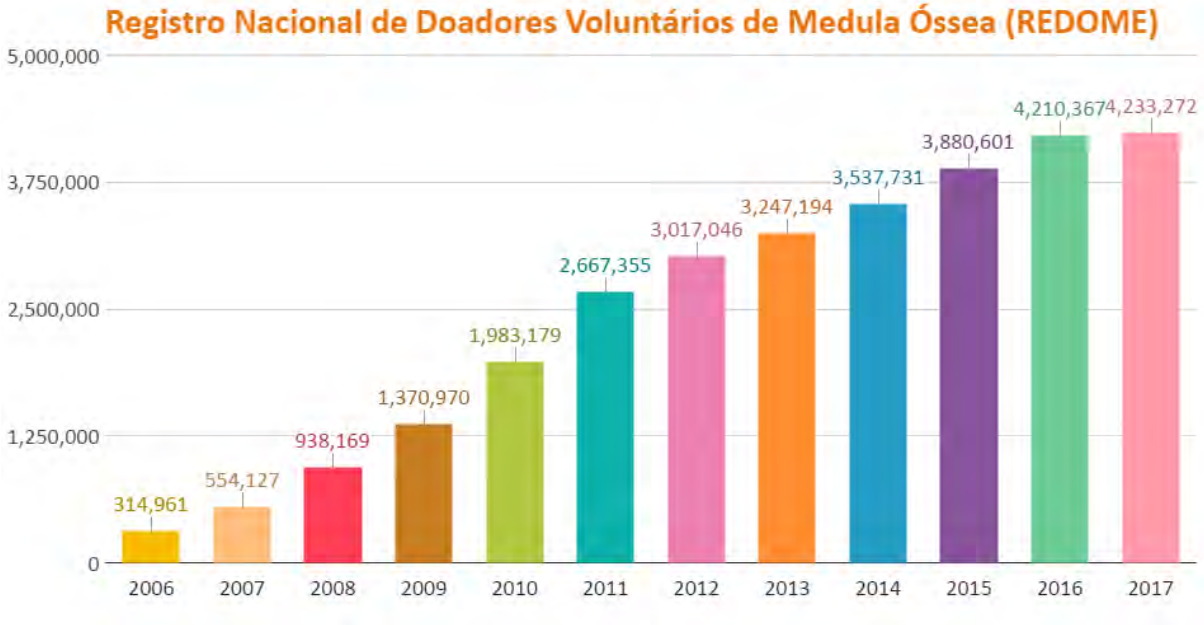
O REDOME atua articulado aos cadastros de todo o mundo. Atualmente, a busca por doadores para pacientes brasileiros é realizada simultaneamente no Brasil e no exterior. Os bancos internacionais também acessam os dados dos candidatos a doadores a partir de sistemas especializados. O registro brasileiro foi o que mais cresceu na última década. A chance de se identificar um doador compatível, no Brasil, na fase preliminar da busca é de até 88%, e ao final do processo, 64% dos pacientes têm um doador compatível confirmado[43].

O Registro Nacional de Doadores Voluntários de Medula Óssea (REDOME) foi criado em 1993, em São Paulo, para reunir informações de pessoas dispostas a doar medula óssea para quem precisa de transplante. Desde 1998, é coordenado pelo Instituto Nacional de Câncer José Alencar Gomes da Silva (INCA), no Rio de Janeiro[42].

A Figura 8 apresenta a evolução do número de doadores cadastrados por ano no REDOME até janeiro de 2017. Também até janeiro de 2017, o número de doadores cadastrados no REDOME era de 4.233.272 e o número de pacientes em busca de doador não aparentado era de 850 [44].

As principais doenças que podem ser tratadas com o transplante de Medula Óssea são: leucemia, anemia aplástica, anemia de fanconi, síndromes provocadas pela deficiência medular (síndromes mielodisplásticas, linfomas, meloma múltiplo, hemoglobinopatias, tumores de testículos e neuroblastomas [45].

Figura 8 - Evolução do número de doadores cadastrados no REDOME por ano até janeiro de 2017



Fonte: [44].

3 MEDICINA LABORATORIAL

Este capítulo aborda a questão dos laboratórios diagnósticos, assim como as vantagens da informatização e a importância do mapeamentos dos processos. É dividido nas seguintes seções: 3.1 Laboratórios diagnósticos, 3.2 Vantagens da tecnologia da informação e 3.3 A importância do mapeamento de processos.

3.1 Laboratórios diagnósticos

Com a descoberta dos grupos sanguíneos e o desenvolvimento da análise química dos fluidos orgânicos na década de 1920, a demanda por testes de laboratório pela comunidade médica cresceu tanto que os hospitais precisaram manter os laboratórios médicos funcionando em tempo integral. Como os patologistas eram os responsáveis por realizar as autópsias, eles naturalmente foram escolhidos para assumir a responsabilidade pelo trabalho centralizado do laboratório. Com o aumento da dependência dos testes de laboratório, que se deu na década de 1960, a área de patologia clinica ganhou grande destaque, o que possibilitou que os clínicos patologistas desenvolvessem suas habilidades, principalmente nas áreas de consultoria médica e de gestão de recursos laboratoriais [46].

Entre as décadas de 1970 e 1990, com a ajuda da automação, de sistemas computacionais e de técnicas de imunoensaio e caracterização molecular, a quantidade de testes realizados anualmente em laboratórios americanos cresceu a uma taxa anual de mais de 12% [46].

Em 1999, a Federação Internacional de Química Clínica (IFCC) em conjunto com o Associação Mundial de Sociedades de Patologia e Medicina Laboratorial (WASPaLM) divulgaram os Princípios da Acreditação de Laboratórios Clínicos. O documento defende que, em nome dos interesses dos pacientes, da sociedade e dos governos, os laboratórios clínicos operem dentro de altos padrões de competência profissional e técnica pelos seguintes motivos [47]:

- As decisões sobre os diagnósticos, prognósticos e tratamentos são frequentemente, baseados nos resultados e interpretações dos exames laboratoriais e danos irreversíveis podem ser causados por resultados errôneos;
- Os usuários dos serviços de laboratórios clínicos (tanto pacientes quanto médicos)

podem não possuir conhecimentos técnicos suficientes para avaliar se um laboratório está operando em um nível satisfatório;

- Os pacientes e, em menor extensão os médicos, podem não ter escolha sobre o laboratório a ser utilizado;
- Os exames laboratoriais podem ser custosos, sendo assim, os pacientes, as seguradoras e o governo, que pagam pelos exames, esperam que o laboratório forneça informações válidas;
- É de interesse dos laboratórios qualificados que sua competência seja verificada através de um processo de inspeção, de comparação com padrões apropriados e afirmação pública de suas boas condições.

Acreditação é uma auditoria externa da capacidade de um laboratório fornecer um serviço de alta qualidade. Isso requer que o laboratório apresente informações que demonstrem [47]:

- Conformidade com os padrões de acreditação publicados;
- Existência de um sistema de gerenciamento que aborde medidas internas e externas de qualidade e que estejam descritas em um manual de gestão da qualidade;
- Um laudo especializado realizado por uma organização de acreditação;

No Brasil, a acreditação não é obrigatória. Segundo a Sociedade Brasileira de Patologia Clínica e Medicina Laboratorial (SBPC/ML) "o laboratório pode escolher o órgão acreditador, baseando-se em sua credibilidade, experiência e conhecimento técnico de seus auditores. Estes incentivam a equipe do laboratório, a alta direção e gerência a alcançarem níveis cada vez mais elevados de qualidade e melhoria contínua".

Alguns órgãos nacionais podem ser citados como exemplo, o Programa de Acreditação de Laboratórios Clínicos (PALC) da Sociedade Brasileira de Patologia Clínica/Medicina Laboratorial (SBPC/ML) e a Organização Nacional de Acreditação (ONA) [48].

No ambiente acadêmico, a medicina laboratorial se desenvolveu utilizando a pesquisa básica e aplicada como base para a prática laboratorial, tendo contribuição das áreas da medicina e da pediatria, assim como as da bioquímica e da microbiologia [46].

Hoje, a medicina laboratorial participa em 70% das decisões clínicas, representando apenas 10% dos custos totais com saúde e as tendências ainda apontam um crescimento na utilização dos exames laboratoriais, principalmente daqueles direcionados à prevenção de doenças, o que demonstra a grande importância dessa especialidade médica no campo da saúde [48].

Como foi colocado, a medicina laboratorial teve grande evolução nos últimos anos. Porém, é uma área que tem sofrido a pressão de prover um serviço de extrema qualidade por um custo cada vez mais baixo [46].

Os serviços da medicina laboratorial incluem diversas ações que envolvem solicitação de exames, coleta de amostras, preparação de amostras, identificação de amostras, transporte e armazenamento de material biológico, processamento e análise de amostras, juntamente a subsequente interpretação e elaboração de relatórios e pareceres. Todas essas ações são essenciais no atendimento ao paciente [49].

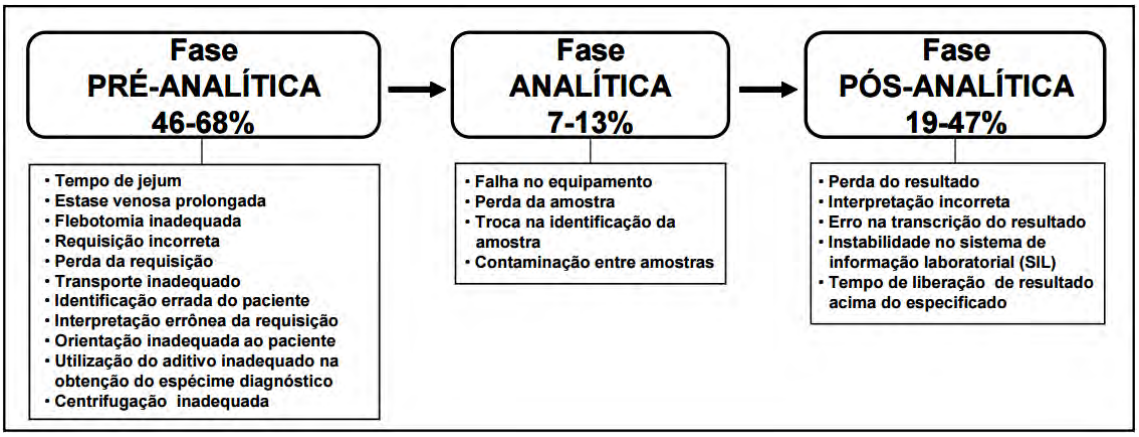
São três as fases do fluxo de processos de um laboratório clínico [48]: pré-analítica, analítica e pós-analítica. A fase pré-analítica compreende os processos anteriores à realização do exame laboratorial de fato, correspondendo à coleta, à manipulação, o processamento e a entrega das amostras aos analisadores. Na fase analítica é realizada a análise do material coletado. A fase pós-analítica compreende o armazenamento das amostras [48].

A Figura 9 apresenta as frequências de erros em cada fase analítica e as referidas variáveis. A fase pré-analítica é a fase que concentra a maior frequência de erros associados a ensaios laboratoriais, entre 46% e 68%. Dentre as variáveis, podem ser citadas o transporte inadequado, a identificação errada do paciente e a centrifugação inadequada do material. [50].

Segundo Lima-Oliveira et al. [50], "o elemento mais sensível na produção de erros na fase pré-analítica diz respeito à atividade humana, em que múltiplos indivíduos interagem no processamento do espécime diagnóstico."

Problemas com o armazenamento e recuperação de dados, com o controle da segurança e com o uso apropriado dos equipamentos e aplicativos do laboratório por pessoal autorizado e treinado podem dificultar o gerenciamento das atividades, além disso, as informações médicas armazenadas nos sistemas de computação podem ser perdidas ou modificadas por pessoal não autorizado [49].

Figura 9 - Fontes e frequências de erro no processamento do espécime diagnóstico



Fonte: [50], f. 442.

3.2 Vantagens da tecnologia da informação

A automatização dos fluxos de trabalho laboratorial pode oferecer ganho de produtividade, redução do custo e facilidade na integração com equipamentos mais modernos. Além disso o uso de sistemas de computação possibilita ao laboratório atingir máxima eficiência e aumentar significativamente a cooperação entre os médicos e reduzir os erros humanos [49].

Segundo Campana et al. [51], "a palavra automação é proveniente do latim *automatus* e significa mover-se por si. Diversas definições estão disponíveis, mas, em resumo, automação é a aplicação de técnicas computadorizadas ou mecânicas com o objetivo de tornar um processo mais eficiente, maximizando a produção com menor gasto de energia e gerando maior segurança."

De modo similar, automação laboratorial consiste na integração entre *hardware* e *software* desenhados para o processamento e análise por completo de amostras. Os métodos automáticos foram introduzidos na medicina laboratorial na década de 1950 pela Technicon por meio dos conceitos de fluxo contínuo [48].

Minimizar erros e reduzir prazos de resultados traz segurança para o paciente, o que gera impacto positivo na economia da saúde [51].

A evolução da automação laboratorial, em todas as suas fases, está diretamente ligada a evolução em tecnologia da informação (TI), incluindo os processos automáticos

pós-analíticos, através do uso de fluxogramas de interpretação, protocolos de repetição e diluição de exames, testes reflexos, entre outros [48].

"A TI é uma das principais forças de mudança no mercado de medicina diagnóstica e permitirá uma integração cada vez maior dos diferentes *players* e segmentos da cadeia, como fontes pagadoras, prestadores, fornecedores etc."[48]. Os sistemas de tecnologia da informação também fornecem procedimentos confiáveis e padronizados para a avaliação de laboratórios médicos [49].

Sistemas de computação são aqueles que automatizam ou apoiam a realização de atividades humanas através do processamento de informações.

Friedman [52] relaciona cinco razões para a TI ser hoje, a principal impulsionadora de toda a indústria de medicina laboratorial:

- A geração de informação aumentará cada vez mais com a massificação dos exames laboratoriais;
- Os desafios do gerenciamento da informação aumentarão à medida que os avanços na genômica e na proteômica gerarem dados cada vez mais complexos e abundantes;
- Os principais desafios profissionais enfrentados por profissionais de laboratório hospitalar são informações de acesso, armazenamento, relatórios e integração de várias fontes;
- As soluções para esses desafios profissionais virão, quase certamente, do setor de TI e não da indústria de diagnósticos *in vitro*;
- As soluções de TI da indústria de diagnósticos *in vitro* são normalmente projetadas para aprimorar a utilidade e a qualidade dos instrumentos e reagentes, ao invés de abordar os desafios do gerenciamento da informação a nível empresarial;

Os laboratórios de hoje e do amanhã são e serão altamente dependentes da TI. A informática é e será utilizada para diferenciar os serviços laboratoriais no mercado. Nessa perspectiva, as principais tendências da tecnologia da informação para o mercado de medicina diagnóstica são [52]:

- Integração técnica e organizacional dos processos dos laboratórios departamentais e gerenciamento da informação em organizações maiores;

- Evolução dos laboratórios clínicos para uma organização digital e/ou virtual designada a servir a redes descentralizadas;
- Evolução dos sistemas de informação laboratorial para sistemas com base em arquitetura baseada em *internet*, com potencial utilização do modelo de linguagem de programação ASP;
- Aprimoramento dos relatórios laboratoriais baseados em texto para relatórios laboratoriais baseados em *internet* e com imagem melhorada;
- Nova ênfase na necessidade de informação dos pacientes e médicos que utilizam os laboratórios, estimulada pelo aumento da sofisticação em TI e do consumismo;
- Aumento da vigilância das questões relativas à segurança e confidencialidade, impulsionadas pela legislação vigente.

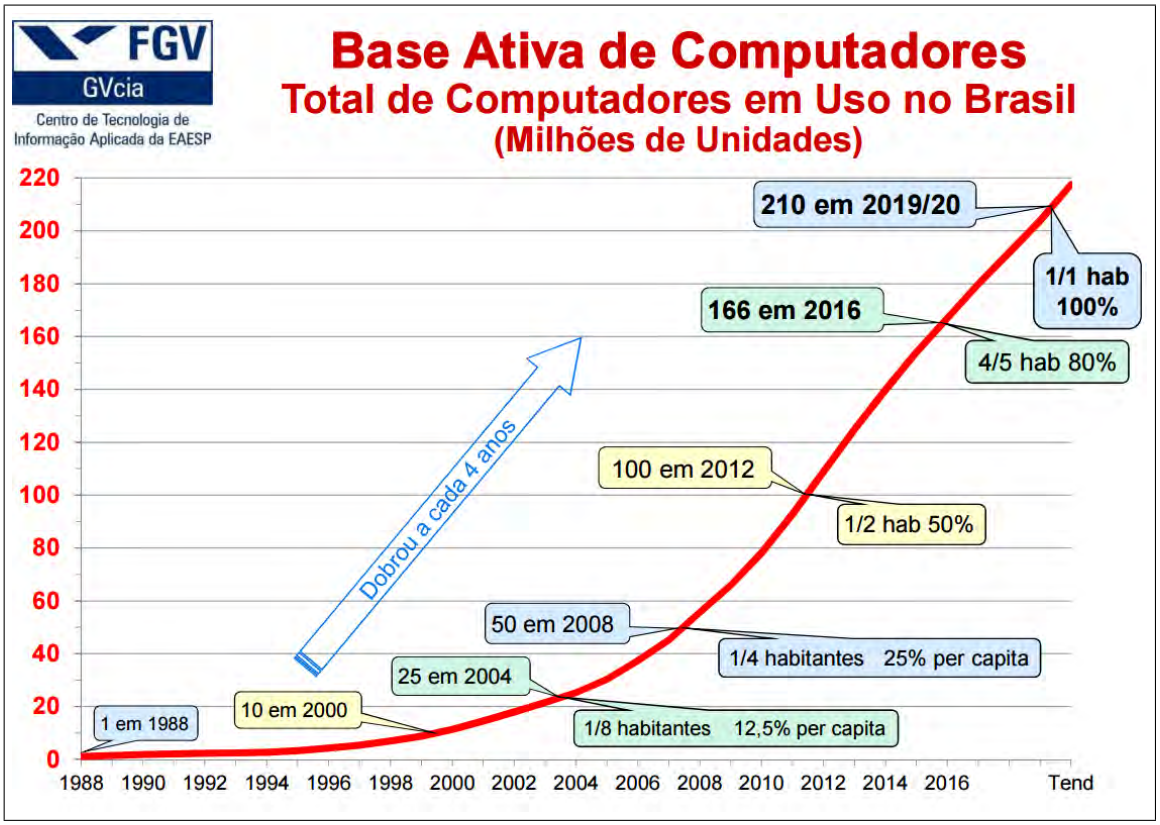
De acordo com a 27ª pesquisa anual, de 2016, realizada pelo Centro de Tecnologia de Informação Aplicada (GVcia) da Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas (FGV/EAESP), que retrata cenário atual e o futuro do mercado de TI, o processo de informatização das empresas e da sociedade ainda é crescente mesmo em uma conjuntura muito adversa e com queda acentuada no consumo pessoal [53].

Apesar de a pesquisa ter indicado queda de 30% em 2015 na venda anual de computadores pela primeira vez na história, a FGV estima crescimento de 7% para 2016. Além disso, o Brasil continua acima da média mundial no uso de computadores por habitante [53].

Segundo a pesquisa da FGV, ao final do ano de 2016 serão 166 milhões de computadores, incluindo *desktops*, *notebooks* e *tablets*, em uso no Brasil, sendo 4 computadores para cada 5 habitantes, o que condiz com 80% *per capita*. Destes 166, 30 milhões corresponderão a *tablets*. Para 2019, a pesquisa aponta o uso de 210 milhões de computadores, sendo um por habitante. [53]. A Figura 10 apresenta o quantitativo de computadores em uso no Brasil através dos anos, desde 1988 até 2016, com previsão para 2019.

A pesquisa ainda assinala que já são 244 milhões de dispositivos móveis conectáveis à *internet* no Brasil, incluindo *notebooks*, *tablets* e *smartphones*. Esse dado corresponde a 1,2 dispositivos portáteis *wireless* por habitante, ou seja, são mais *smartphones* do que computadores ou 6 *smartphones* para cada *tablet* [53].

Figura 10 - Total de computadores em uso no Brasil



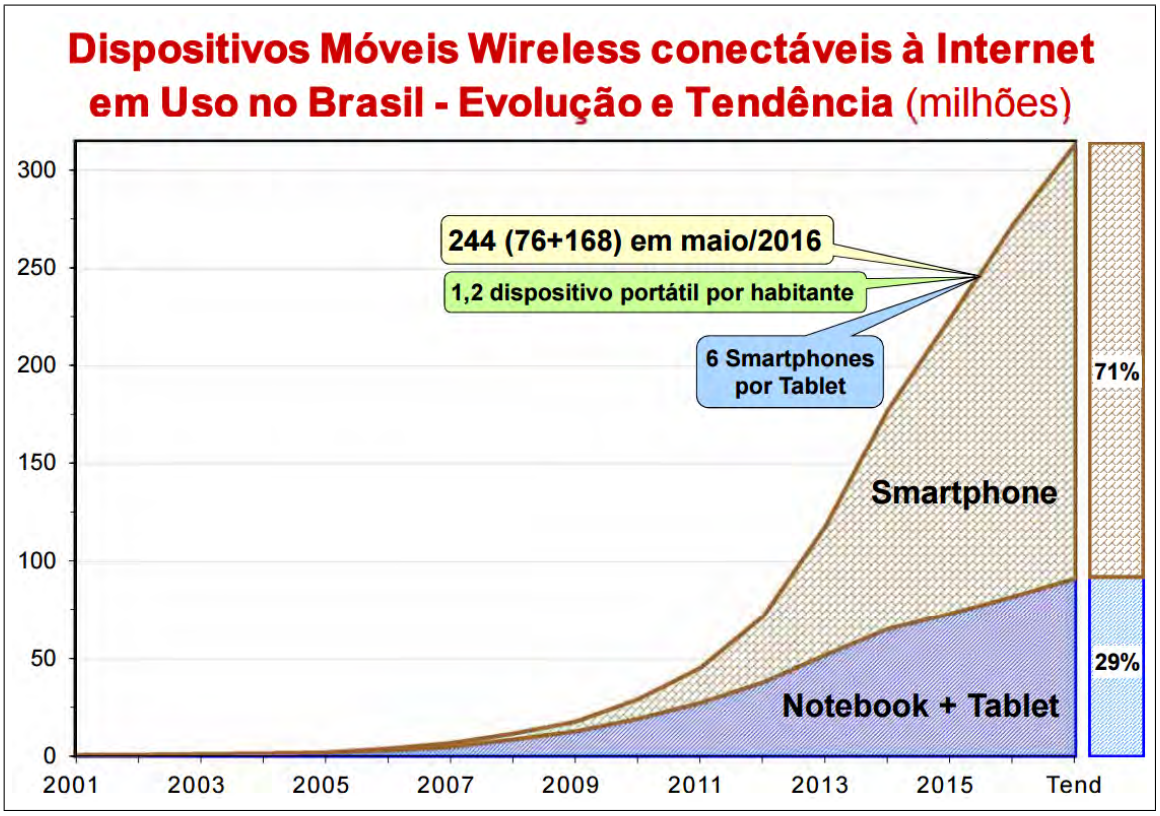
Fonte: [53].

A Figura 11 apresenta a evolução quantitativa de dispositivos móveis *wireless* conectáveis à *internet* em uso no Brasil.

Convém ressaltar que o gasto e investimento em TI ficou estável em 7,6% da receita das empresas. Nas indústrias de capital aberto, durante a última década, para cada 1% a mais de gasto e investimento em TI, depois de 2 anos, o lucro aumentou 7% [53].

Diante dessa realidade, não se pode ignorar o impacto que a TI representa na sociedade. É consenso quase geral que a informatização de processos só traz benefícios para as organizações. Dentre as vantagens destacam-se maior agilidade, segurança e qualidade técnica para as atividades, menor tempo de processamento de lotes e de correção do processo produtivo, menor índice de retrabalhos e erros, o que consequentemente, reduz o tempo de entrega [54].

Figura 11 - Dispositivos móveis *wireless* conectáveis à *internet* em uso no Brasil



Fonte: [53].

A rastreabilidade total dos processos e a facilidade de acesso às informações com maior confiabilidade operacional, a redução do trabalho administrativo e a melhor gestão do laboratório por meio da automação e gestão das atividades diárias, a facilidade no cumprimento e atendimento aos requisitos de integridade, confidencialidade e autenticidade dos registros também são elementos que a informatização agrega aos processos [54].

3.3 A importância do mapeamento de processos

Mapeamento de processos é uma ferramenta analítica e de comunicação destinada a ajudar a melhorar os processos existentes ou a implementar uma nova estrutura de processos. É uma excelente ferramenta gerencial que pode ser utilizada para melhor entender o processo atual e eliminar ou simplificar aqueles que necessitam de mudança. Um processo de negócio é uma série de etapas projetadas para produzir um produto ou um serviço [55], sendo assim, todo e qualquer serviço prestado a um cliente necessita de

um processo para sua realização.

A organização não pode correr o risco de perder lições e experiências conseguidas ao longo de muitos anos em função de seus funcionários mudarem de emprego ou se aposentarem, já que o aprendizado é construído com base na troca de conhecimentos e experiências entre os indivíduos [11].

Pode-se dizer, então, que o mapeamento de processos desempenha o papel essencial de desafiar os processos existentes, criando oportunidades de melhoria de desempenho organizacional ao identificar interfaces críticas e, sobretudo, criar bases para implantação de novas e modernas tecnologias de informação (TI) e de integração empresarial [11].

Em trabalho de laboratório de rotina, processar um grande número de amostras e inserir os resultados manualmente significa que a probabilidade de os resultados ficarem fora dos limites especificados é maior, ou seja, desvios ocorrerão com mais frequência. A investigação de acompanhamento das fontes de erro é demorada, uma vez que todas as fontes de erro possíveis devem ser consideradas.

Eliminando-se a necessidade de operações manuais, dependendo da quantidade de amostras, um tempo considerável poderá ser economizado. Supondo que se leve cerca de 30 segundos para transcrever o resultado de uma amostra para uma tabela de forma manual e que de forma automática leve menos de 1 segundo, em um laboratório que processa 500 amostras por dia, essa economia pode ser de aproximadamente 250 minutos por dia, o que corresponderia a um pouco mais de 4 horas.

Observada a relevância de se conhecer tais processos produtivos, são utilizadas técnicas de modelagem de processos de negócio para representar as diversas tarefas necessárias, na sequência em que elas ocorrem para a realização e entrega do serviço. O mapa de processos deve ser apresentado graficamente, permitindo expor os detalhes do processo, encorajar concisão e precisão na sua descrição, focar a atenção nas interfaces do mapa do processo e fornecer uma análise de processos consistente com a linguagem do projeto [11].

São quatro as fase de elaboração de um mapa de processos [55]:

1) **Fase de coleta de dados:** envolve atividades de pesquisa da área de estudo e entrevistas com os que estão mais familiarizados com o processo;

2) **Fase de estruturação do mapa:** envolve atividades de criação do mapa através da elaboração de rascunhos utilizando edição gráfica, com o objetivo de esclarecer,

simplificar e rearranjar as informações coletadas na fase anterior;

3) **Fase de documentação do mapa:** envolve atividades de aperfeiçoamento, revisão e correção do mapa criado na fase anterior e de elaboração de texto devidamente estruturado;

4) **Fase de interação de *feedback*:** envolve a apresentação do mapa elaborado para considerações e comentários, o que pode resultar em nova revisão do mapa.

4 PROCESSO DE TIPIFICAÇÃO HLA DO HLA-UERJ

Este capítulo traz uma apresentação do laboratório de Histocompatibilidade e Criopreservação (HLA-UERJ) da Universidade do Estado do Rio de Janeiro, detalha o processo de tipificação HLA do HLA-UERJ atual e aponta oportunidades de melhoria e os recursos necessários à implementação. Está dividido nas seguintes seções: 4.1 Caracterização do ambiente de estudo, 4.2 Detalhamento e mapeamento do processo atual e 4.3 Identificação de oportunidades de melhoria.

4.1 Caracterização do ambiente de estudo

O estudo foi ambientado junto ao Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro (HLA-UERJ), com os objetivos principais de dirimir riscos operacionais e ampliar a capacidade de análise e rastreabilidade dos dados gerados.

4.1.1 Laboratório HLA-UERJ

O Laboratório de Histocompatibilidade e Criopreservação da Universidade do Estado do Rio de Janeiro (HLA-UERJ) teve início no final do ano de 1994 e a princípio ocupou algumas salas do Departamento de Histologia e Embriologia do Instituto de Biologia da UERJ. [6].

Em 1996, o HLA-UERJ foi credenciado pela Associação Brasileira de Histocompatibilidade (ABH) e passou a ser responsável pelos exames de imunogenética relacionados aos transplantes renais realizados no Hospital Universitário Pedro Ernesto (HUPE). Além disso, passou a ser colaborador no cadastramento de doadores voluntários do Registro Nacional de Doadores de Medula Óssea (REDOME). Em 2003, o laboratório foi transferido para a Policlínica Piquet Carneiro e em 2007, estendeu suas atividades para o uso de células e tecidos preservados [6].

Atualmente, o HLA-UERJ está localizado na Policlínica Piquet Carneiro e atua entre as áreas de assistência, ensino e pesquisa. Sua equipe é composta por 25 colaboradores entre servidores, prestadores de serviço, pesquisadores e alunos de pós-graduação. Suas atividades envolvem a realização de exames laboratoriais na área de compatibilidade

para transplante de órgãos sólidos, exame moleculares referentes às hepatites virais (B e C), criopreservação de células-tronco hematopoéticas, participação em campanhas para cadastramento de doadores voluntários de medula óssea, oferecimento de cursos de formação e aperfeiçoamento, principalmente treinamento em histocompatibilidade e biologia molecular e participação em projetos de pesquisa [6].

O HLA-UERJ possui acreditação do Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), concedida com base no cumprimento dos requisitos das normas e em documentos normativos [6].

4.1.2 Atividades do Laboratório HLA-UERJ

Desde o seu início, o laboratório HLA-UERJ se dedica a estudos de associação de alelos do complexo principal de histocompatibilidade com doenças, priorizando as áreas de reumatologia e dermatologia. Outra área de pesquisa na qual o laboratório tem contribuído é a de transplantes renais, ajudando na detecção e prevenção de rejeições crônicas através de exames laboratoriais que permitam acompanhar a evolução de pacientes no pós-transplante, o que pode determinar o tratamento mais eficaz para o paciente [6].

Na atualidade podemos relacionar as atividades prestadas pelo HLA-UERJ como sendo as seguintes [6]:

- **Tipificação HLA:**
  - Identificação de doador aparentado de células tronco (1ª, 2ª e 3ª fases);
  - Identificação de doador não aparentado de células tronco (1ª, 2ª e 3ª fases);
  - Identificação de receptor de células tronco (1ª, 2ª e 3ª fases).
- **Provas Cruzadas:**
  - Provas cruzadas em doador vivo de rim;
  - Auto prova cruzada em doador vivo de rim;
  - Prova cruzada em doador vivo de rim contra linfócitos T ou B;
  - Identificação de doador vivo de rim (1ª e 2ª fases);
  - Avaliação de reatividade do receptor de rim (1ª fase);
  - Identificação de receptor de rim (1ª fase).

- **Detecção de RNA do vírus da Hepatite C (Qualitativo)**;
- **Quantificação de RNA do vírus da Hepatite C (Carga Viral)**;
- **Genotipagem de vírus da Hepatite C**.

4.1.3 Exames de Tipificação HLA do HLA-UERJ

O Laboratório HLA-UERJ realiza diversos exames para identificar a tipagem HLA com o intuito de avaliar a compatibilidade entre doadores e receptores, contribuindo para o sucesso de transplantes de órgão e células-tronco hematopoéticas.

Abaixo seguem alguns dos exames de tipificação HLA praticados pelo laboratório HLA-UERJ [6]:

- **Tipificação HLA de Doador Aparentado de células-tronco hematopoéticas:** Este exame permite a tipificação HLA com a identificação dos alelos HLA-A, HLA-B e HLA-DRB1 no nível de baixa ou média resolução dos candidatos à doação aparentada de células-tronco hematopoéticas e deverá ser solicitado pelo médico hematologista responsável. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias;
- **Tipificação HLA de Doador Voluntário de células-tronco hematopoéticas – 1ª fase:** Este exame permite a tipificação HLA com a identificação dos alelos HLA-A, HLA-B e HLA-DRB1, no nível de baixa resolução dos doadores voluntários cadastrados pelo REDOME e deverá ser solicitado pelo hemocentro responsável. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias;
- **Tipificação HLA de Doador Voluntário de células-tronco hematopoéticas – 2ª fase:** Este exame permite a complementação da tipificação HLA de doadores voluntários, potencialmente compatíveis com pacientes candidatos a transplante de células-tronco hematopoéticas. Consiste na identificação dos alelos HLA-C e HLA-DRB1, no nível de alta resolução de doadores voluntários previamente selecionados e deverá ser solicitado pelo REDOME. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias;

- **Tipificação HLA Confirmatória de Doador Voluntário de células-tronco hematopoéticas:** Este exame permite a confirmação da tipificação HLA de doadores voluntários candidatos ao transplante de céulas-tronco hematopoéticas; consiste na identificação dos alelos HLA-A, HLA-B, HLA-C, HLA-DRB1 e HLA-DQB1 no nível de alta resolução, e deverá ser solicitado pelo REDOME. Este exame faz uso das metodologias PCR-SSO e sequenciamento de DNA e possui prazo de 12 dias;

- **Tipificação HLA de Receptor de transplante de células-tronco hematopoéticas:** Este exame permite a tipificação HLA com a identificação dos alelos HLA-A, HLA-B e HLA-C, no nível de baixa ou média resolução, e HLA-DRB1 e HLA-DQB1 em alta resolução dos pacientes candidatos a transplante de células-tronco hematopoéticas e deverá ser solicitado pelo médico hematologista responsável. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias;

- **Tipificação HLA Confirmatória de Receptor de células-tronco hematopoéticas:** Este exame permite a confirmação da tipificação HLA de pacientes candidatos ao transplante de céulas-tronco hematopoéticas. Consiste na identificação dos alelos HLA-A, HLA-B, HLA-C, HLA-DRB1 e HLA-DQB1 no nível de alta resolução e deverá ser solicitado pelo REDOME. Este exame faz uso das metodologias PCR-SSO e sequenciamento de DNA e possui prazo de 12 dias;

- **Tipificação HLA de Receptor de Rim:** Este exame permite a tipificação HLA com a identificação dos alelos HLA-A, HLA-B e HLA-DRB1, no nível de baixa ou média resolução dos candidatos a transplante renal e deverá ser solicitado pelo médico nefrologista responsável. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias;

- **Tipificação HLA de Doador Vivo de Rim – 1ª fase:** Este exame permite a tipificação HLA com a identificação dos alelos HLA-A, HLA-B e HLA-DRB1 no nível de baixa ou média resolução dos candidatos a doação para transplante renal e deverá ser solicitado pelo médico nefrologista responsável. Este exame faz uso da metodologia PCR-SSO e possui prazo de 8 dias.

4.2 Detalhamento e mapeamento do processo atual

Para que o detalhamento e mapeamento do processo atual de tipificação HLA fosse realizado, contou-se com a colaboração dos funcionários do HLA-UERJ, uma vez que eles são os principais atuantes do processo. Com base em informações coletadas no local, foi possível encadear as atividades e mapear o processo atual. As entrevistas foram realizadas durante o segundo semestre do ano de 2016 e o primeiro semestre de 2017.

No processo de tipificação HLA, podem ser utilizados métodos de baixa, média e alta resolução. Na tipagem de baixa resolução, somente os dois primeiros dígitos são identificados, indicando a que grupo os alelos pertencem, grupos esses que, em geral, correspondem às especificidades sorológicas: A*03 (A3), B*07 (B7), DRB1*03 (DR3), etc. A análise de alta resolução, que utiliza um conjunto adicional de *primers*, permite a identificação dos alelos propriamente ditos: A*0302, B*0702, DRB1*0301, etc. [56]. Os métodos que fornecem informações entre o nível sorológico e o nível de alelos são denominados métodos de média resolução.

De forma geral, o processo de tipificação HLA envolve técnicas de Reação em Cadeia da Polimerase seguida de hibridização com Oligonucleotídeos de Sequência Específica (PCR-SSO) (método de média resolução) e técnicas de sequenciamento de DNA utilizando o método de Sanger e o método de nova geração (NGS) (método de alta resolução). Como o método de NGS é menos custoso do que o método de Sanger, o laboratório tem o objetivo de substituir por completo o método de Sanger pelo o método de NGS, portanto o método de Sanger não será incluído neste mapeamento.

As amostras de cada doador ou paciente são recebidas de forma individualizada e em tubos, que são dispostos em uma centrífuga. A centrífuga é um equipamento que realiza a função de separar as células que são detentoras de DNA [57], pois em uma amostra as células estão todas juntas e misturadas, como hemácias, leucócitos, plaquetas, entre outras. A centrífuga age de acordo com o gradiente de densidade de cada uma das células, já que devido ao movimento de rotação, o processo de decantação é acelerado e as partículas de maior densidade são arremessadas para o fundo dos tubos. Sendo assim, após a centrifugação, é possível notar três camadas distintas no tubo: uma camada superficial de plasma, que é a parte líquida do sangue, uma camada de leucócitos que é mais esbranquiçada, conhecida como "creme leucocitário"e uma camada de hemácias compactadas no fundo. Como as hemácias não possuem núcleo, sendo, portanto, desprovidas de DNA, a

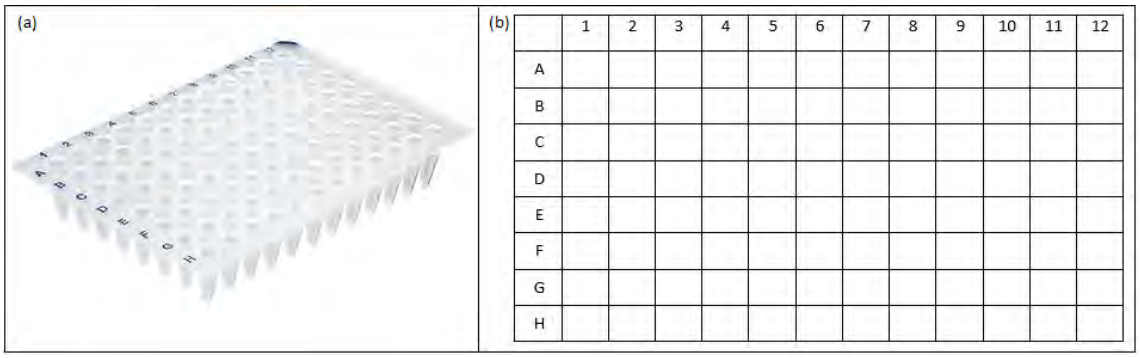
camada alvo do processo de tipificação HLA é a que possui o "creme leucocitário", sendo os leucócitos os elementos utilizados como substrato para a extração do DNA.

Com a centrifugação concluída, cerca de 200 microlitros do "creme leucocitário"de cada amostra são utilizados para a extração de DNA de fato. O restante é armazenado para que o laboratório possa ter a opção de voltar à amostra se assim desejar, que nesse ponto é chamada de amostra "estoque".

Cada paciente ou doador recebe um código interno de identificação, chamado de "número de DNA". Esse código é um número sequencial único. Em uma planilha Excel (Microsoft® Office™) fica armazenada a listagem dos pacientes e doadores com informações tais como nome do paciente ou doador, sua origem e o código único atribuído.

O próximo passo é realizar a amplificação do DNA extraído através da técnica PCR-SSO e para tal, é utilizada uma microplaca com poços onde são dispostas as amostras. A preparação da placa envolve a produção de um mapa de localização e identificação das amostras no estilo de uma tabela, com linhas e colunas, representando exatamente a microplaca. Na Figura 12 são apresentados um exemplo de microplaca com 96 poços e um exemplo do mapa de identificação das amostras na microplaca de PCR-SSO.

Figura 12 - Exemplo de microplaca e seu mapa de identificação. (a) Exemplo de microplaca com 96 poços. (b) Exemplo de um mapa produzido e utilizado pelo laboratório para identificação das amostras na microplaca de PCR-SSO



Fonte: (a) Imagem retirada da *Internet*, 2017. (b) O autor, 2017.

A seleção das amostras é realizada de forma manual, ou seja, as amostras são escolhidas utilizando-se a planilha Excel (Microsoft® Office™) de listagem dos pacientes e doadores como referência. Em seguida, um mapa de localização e identificação das

amostras na microplaca de PCR-SSO em branco é impresso em papel. O "número de DNA" de cada amostra selecionada é transcrito à mão para o mapa, respeitando-se a disposição das amostras na microplaca. Os mapas preenchidos são arquivados em pastas arquivo que são armazenadas no laboratório para consultas posteriores. Nesse passo são 32 amostras analisadas e como para cada amostra são feitas 3 investigações, são 96 poços preenchidos na microplaca.

Com as amostras selecionadas e o mapa preenchido, a microplaca é submetida ao termociclador para a amplificação do DNA pelo método PCR-SSO. Termociclador é um equipamento que eleva e abaixa a temperatura do material de forma programável em vários ciclos [57]. O resultado do termociclador é analisado quanto a sua qualidade utilizando-se da eletroforese em gel e foto documentação. A eletroforese em gel é uma técnica de separação de partículas de acordo com sua carga elétrica e peso molecular utilizando-se um determinado gel como condutor [57].

Após a análise, caso a amostra apresente algum problema ou estiver fora do esperado, a mesma é descartada e retirada da placa, não seguindo para o próximo passo. Nesse caso, o mapa das amostras na microplaca deve ser refeito eliminando-se a marcação do campo correspondente ao material descartado, porém deve ser mantida a marcação das amostras aprovadas, ou seja, a placa não é rearranjada. A Figura 13 apresenta um exemplo de um mapa produzido e utilizado pelo laboratório para identificação das amostras na microplaca antes de ser submetida ao termociclador e um exemplo de um mapa de identificação das amostras na microplaca refeito após as amostras serem submetidas ao termociclador e algumas serem descartadas.

As amostras aprovadas são encaminhadas para o próximo passo. A metodologia seguinte é escolhida de acordo com o pedido do médico ou do hemocentro solicitante da análise, podendo ser tanto de média resolução quanto de alta resolução.

Figura 13 - Exemplo de mapa de localização e identificação das amostras na microplaca. (a) Exemplo de um mapa produzido e utilizado pelo laboratório para identificação das amostras na microplaca antes de ser submetida ao termociclador. (b) Exemplo de um mapa de identificação das amostras na microplaca refeito após as amostras serem submetidas ao termociclador e algumas serem descartadas

(a)

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	1	1	2	2	2	3	3	3	4	4	4
B	5	5	5	6	6	6	7	7	7	8	8	8
C	9	9	9	10	10	10	11	11	11	12	12	12
D	13	13	13	14	14	14	15	15	15	16	16	16
E	17	17	17	18	18	18	19	19	19	20	20	20
F	21	21	21	22	22	22	23	23	23	24	24	24
G	25	25	25	26	26	26	27	27	27	28	28	28
H	29	29	29	30	30	30	31	31	31	32	32	32

(b)

	1	2	3	4	5	6	7	8	9	10	11	12
A	1	1	1	2	2	2	3	3	3	4		4
B		5	5	6	6	6	7	7	7	8	8	
C	9	9	9	10		10	11	11	11	12	12	12
D	13	13	13	14	14	14	15	15	15	16	16	16
E	17		17	18	18	18	19		19	20	20	20
F	21	21	21	22	22	22	23	23	23	24	24	24
G	25	25	25		26	26	27	27	27		28	28
H	29	29	29	30	30	30		31	31	32	32	32

Fonte: (a) e (b) O autor, 2017.

Para o procedimento de hibridização do DNA com Oligonucleotídeos de Sequência Específica (SSO), é utilizado um citômetro de fluxo (Luminex®). A citometria de fluxo é um método para o isolamento de células e cromossomos através da utilização de corantes fluorescentes [57], assim, o citômetro de fluxo é capaz de contar, examinar e classificar essas células e permite a identificação da especificidade antigênica. Ex: HLA-A*03 (A3), que corresponde à família de antígenos HLA-A3. O equipamento possui um computador acoplado contendo uma porta USB. Esse computador não está conectado à rede por recomendação do fornecedor, a fim de preservar o sistema contra vírus e outras ameaças. A cada análise executada, é gerado um arquivo de saída no formato *.csv. Todos os arquivos gerados são copiados através da porta USB para um computador do laboratório. Os técnicos utilizam o próprio *software* do fornecedor para analisar os dados.

Para o procedimento de NGS, o preparo das amostras envolve um ajuste de concentração a ser aplicado, conforme especificação do fabricante do *kit* de reagentes a ser utilizado. Esse *kit* de reagentes não acompanha o equipamento de sequenciamento, é adquirido a parte pelo laboratório.

Para agilizar esse ajuste de concentração de cada amostra, já que no NGS são realizadas várias determinações por amostra, às vezes até mais que três, podendo a quantidade de amostras aumentar significativamente chegando ao número de 560 de forma aproximada, foi criada uma planilha Excel (Microsoft® Office™) com uma aba exclusiva, que pode ser chamada de aba de referência, com os dados fornecidos pelo fabricante

do *kit* a respeito dos volumes e concentrações pré-determinados. Em outra aba ficam armazenados os dados dos pacientes ou doadores, que são sigilosos, tais como: identificação do material, nome, quantificação do DNA extraído, entre outros, essa aba seria a aba da lista das amostras.

A quantificação de DNA de cada amostra é realizada por espectrofotometria no equipamento NanoDrop® , que é um equipamento que avalia a quantidade geral de DNA na amostra, sua pureza e integridade [57]. O NanoDrop® realiza a medição e os dados são copiados através de uma porta USB presente no dispositivo. Esses dados alimentam a aba da lista das amostras na planilha.

Sendo assim, é realizada uma comparação automática entre a aba da lista das amostras e a aba de referência, utilizando-se fórmulas de pesquisa e referência pré-definidas na própria planilha para que se obtenha o valor de concentração necessário a ser ajustado em cada amostra para que se atinja o padrão pré-estabelecido pelo fabricante do *kit*.

Por exemplo, se uma amostra que já foi quantificada possui sua concentração acima do determinado pelo fabricante do *kit* de reagentes, a mesma precisará ser diluída em solução aquosa e vice-versa. Como a planilha já foi configurada para realizar os cálculos de forma automática, os valores necessários para o ajuste de concentração já são mostrados para o técnico que está executando o procedimento na aba da lista das amostras.

Com as concentrações corretas, a microplaca é preparada com as amostras e um mapa de localização e identificação das amostras na microplaca de NGS é produzido e impresso, utilizando-se a mesma planilha Excel que auxilia o ajuste de concentração. Diferentemente do mapa de PCR-SSO, o mapa da microplaca de NGS é gerado automaticamente através de fórmulas de referência do próprio Excel.

O equipamento que o laboratório adquiriu para esse passo foi o HiSeq 2500 (Illumina®). O HiSeq 2500 é um sistema de sequenciamento que fornece eficiência para genômica em larga escala capaz de gerar dados com qualidade a uma taxa de 160Gb (gigabytes) por dia ou 1Tb (terabytes) por ciclo de investigação [58]. O sistema compreende o equipamento de sequenciamento, um monitor e um computador de controle com teclado, mouse e leitor de código de barras [59]. A Tabela 3 apresenta as especificações do computador de controle que integra o sistema HiSeq 2500 (Illumina®). O sequenciador permite a identificação dos alelos propriamente ditos, compostos com os 4 dígitos. Ex: HLA-A*0210, que representa o décimo alelo HLA-A2 dentro da família de antígenos

HLA-A2 serologicamente definidos.

O resultado da análise de uma execução de sequenciamento é um conjunto de arquivos de identificação de bases avaliados por qualidade com a extensão *.bcl. Para cada bloco que é analisado, um arquivo comprimido por ciclo é gerado [59]. Entretanto, o formato *.fastq é o mais amplamente utilizado para representar e analisar sequências, já que nesse formato as bases e os valores de qualidade são representados no mesmo arquivo, que são arquivos de texto [60]. Mediante a isso e para que se possa utilizar a maioria das plataformas de tipagem HLA, é necessária que seja realizada a conversão dos arquivos *.bcl para arquivos *.fastq.

Tabela 3 - Especificação técnica do computador de controle do sistema HiSeq 2500 (Illumina®)

Especificação do Computador HiSeq 2500	
CPU	Intel® Xeon® CPU E5-2620 @ 2.00 GHz (2 processadores)
MEMÓRIA	64 GB RAM
DISCOS	8 x 1.0 TB 3.5" SATA (2 OS - RAID1 drives e 6 data RAID5 drives)
SISTEMA OPERACIONAL	Microsoft® Windows® 7 64-bit

Fonte: [58].

Para realizar a conversão dos arquivos *.bcl para *.fastq, o HLA-UERJ utiliza a plataforma de análise genômica integrada ao sistema HiSeq, chamada BaseSpace®. BaseSpace® é um poderoso serviço de computação em nuvem fornecido pela Illumina® onde biólogos e informáticos podem facilmente armazenar, analisar e compartilhar dados genéticos. Para isso, basta que seja criada uma conta de acesso em basespace.illumina.com. Os arquivos *.bcl são enviados através da *Internet* em tempo real para o espaço dedicado na nuvem [61]. Os dados são automaticamente convertidos em *.fastq, gerando dois arquivos comprimidos por amostra, resultando em torno de 100Mb (megabytes) de dados por placa analisada. Quando a conversão é completada, os arquivos são copiados do equipamento HiSeq utilizando-se um disco rígido externo.

Com os arquivos já convertidos e copiados para um computador do laboratório, é iniciado o passo de descompressão dos arquivos. Nesse passo, cada arquivo é descom-

pactado de forma manual por um técnico (caso sejam analisadas 560 amostras, o técnico precisará descompactar 1120 arquivos).

O *software* de interpretação da tipagem HLA utilizado pelo laboratório é o NGSEngine® (GenDx). O NGSEngine® é um pacote de *softwares* pago, independentes de plataforma, desenvolvido para realizar tipagem HLA baseada em dados adquiridos pelo método NGS, capaz de assinalar alelos HLA de forma confiável e inequívoca. De acordo com um dos técnicos do laboratório, com os computadores que o laboratório possui, o NGSEngine® (GenDx) não é capaz de analisar mais do que 96 amostras em média ao mesmo tempo. Acima desse número de amostras, o sistema trava e para de funcionar. A Tabela 4 apresenta as especificações de um dos computadores do HLA-UERJ utilizado para execução do *software* de interpretação da tipagem HLA.

Tabela 4 - Especificação técnica de um computador do HLA-UERJ

Especificação do Computador HLA-UERJ	
CPU	Intel® Core™ CPU i5-3330 @3.00GHz 3.20GHz
MEMÓRIA	8GB GB RAM
SISTEMA OPERACIONAL	Microsoft® Windows® 7 64-bit

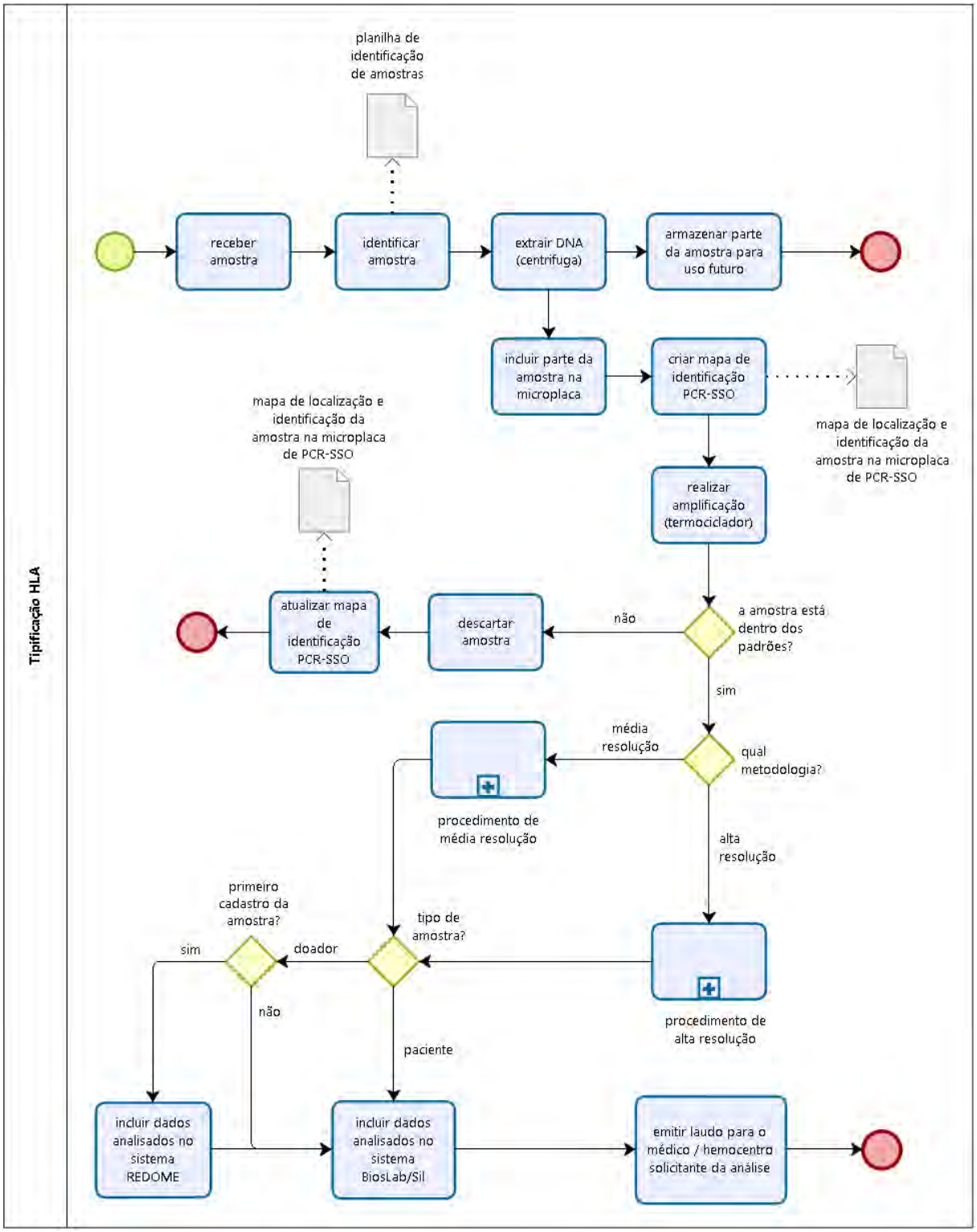
Fonte: O autor, 2017.

Após a análise, seja do método de média resolução, seja do método de alta resolução, é gerado um relatório com os dados que alimenta outros dois sistemas: um é o sistema laboratorial chamado BiosLab/Sil e o outro é o REDOME. O REDOME só é alimentado com os dados da amostra do doador, caso seja o primeiro cadastro da amostra.

Todos os dados envolvidos no processo e analisados devem ser guardados respeitando-se as normas laboratoriais, o que engloba as amostras e os mapas de localização e identificação das amostras nas microplacas.

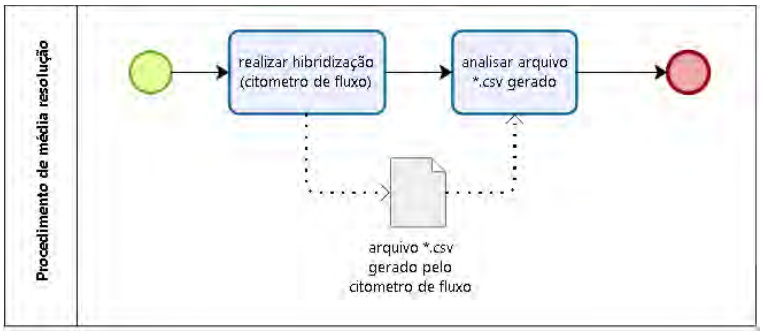
Para o mapeamento do fluxo do processo de tipificação HLA, foi utilizada a técnica de modelagem de processos conhecida por BPMN (*Business Process Model and Notation*), em português Notação de Modelagem de Processos de Negócio, já que o BPMN oferece uma notação padrão largamente aceita [62]. A Figura 14 apresenta o fluxo do processo de tipificação HLA do HLA-UERJ. As Figura 15 e Figura 16 apresentam, respectivamente, os subprocessos de média e alta resolução.

Figura 14 - Fluxo do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler*



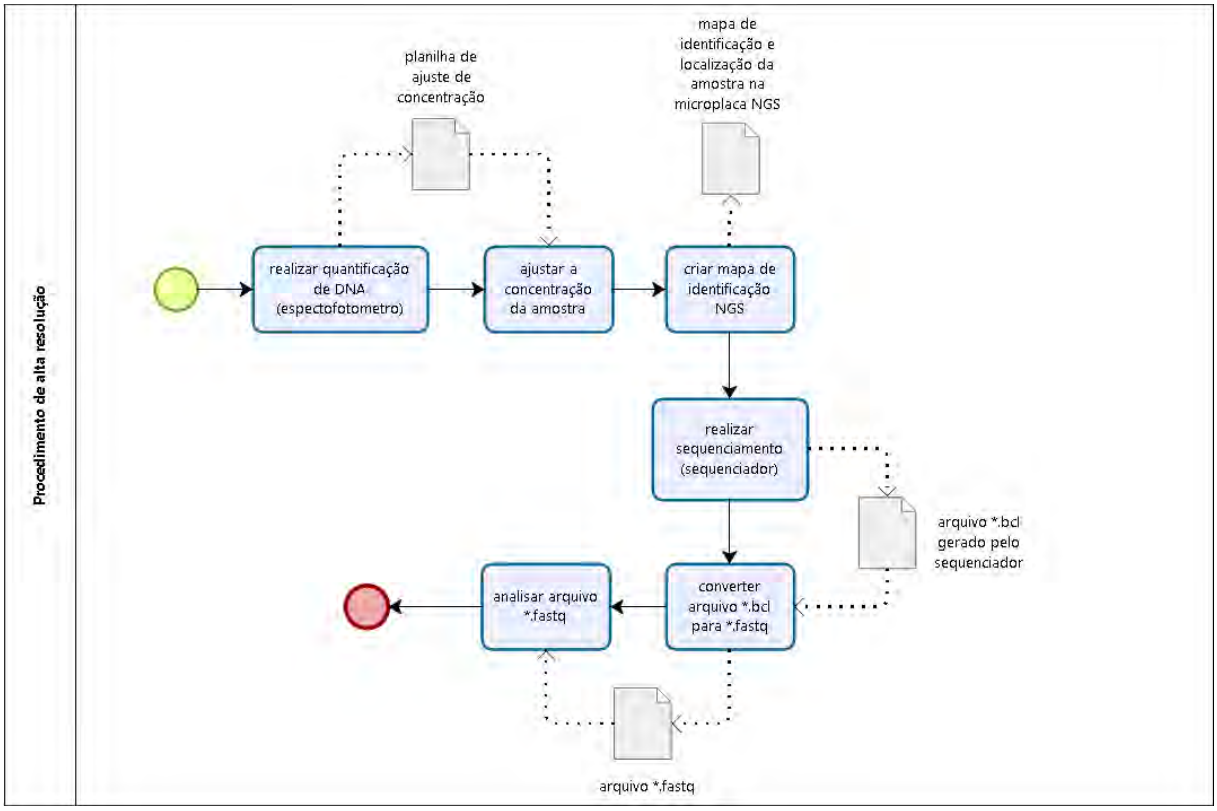
Fonte: O autor, 2017.

Figura 15 - Fluxo do subprocesso de média resolução do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler*



Fonte: O autor, 2017.

Figura 16 - Fluxo do subprocesso de alta resolução do processo de Tipificação HLA do HLA-UERJ, criado na ferramenta *Bizagi Modeler*



Fonte: O autor, 2017.

4.3 Identificação de oportunidades de melhoria

Após a realização do mapeamento das atividades de tipificação HLA, pode-se observar e analisar a ocorrência de alguns riscos e vulnerabilidades no processo atual.

Foi constatada a ausência de uma gerência central do processo mapeado. O controle das atividades é descentralizado e os dados estão espalhados em diferentes arquivos. Essa forma de armazenar as informações prejudica uma possível rápida recuperação, dificulta o controle da redundância e põe em dúvida a integridade dos dados, o que afeta, direta ou indiretamente, a qualidade das atividades.

Como os mapas de localização e identificação das amostras nas microplacas são elaborados em papel e preenchidos à mão, eles possuem limitações consideráveis. Uma delas é que o papel ocupa muito espaço físico e são geralmente, armazenados em pastas arquivo guardadas em armários. Portanto, qualquer consulta aos dados torna-se extremamente difícil, demorada e inadequada. Outra limitação é a caligrafia, que pode ser confusa ou até ilegível. Sem falar que o documento pode ser perdido ou danificado acidentalmente, ou até mesmo de forma maliciosa.

Muitas das limitações mencionadas podem ser reduzidas através da implantação de um sistema de tecnologia da informação que controle o fluxo de trabalho e centralize todos os dados utilizados no processo, o que pode proporcionar maior acurácia, precisão, acessibilidade, rastreabilidade e confiabilidade dos dados.

Os sistemas de controle de *workflow* ou sistemas de gerenciamento de *workflow* (SGWs) facilitam a automação, o gerenciamento de processos e controlam a circulação das atividades. São capazes de armazenar documentos em um local central, conseguindo apresentar informações em formato unificado, melhorando o trabalho em equipe, fornecendo suporte automatizado de processo e permitindo que os membros da equipe se comuniquem e colaborem dentro de um ambiente padronizado [63].

Com isso, algumas vantagens do uso de SGWs podem ser destacadas [64]:

- **Maior eficiência:** a automação de muitos processos de negócio resulta na eliminação de muitas etapas desnecessárias;
- **Melhor controle do processo:** um melhor gerenciamento dos processos de negócio é atingido através da padronização dos métodos de trabalho e da disponibilidade de registros de auditoria;

- **Melhor atendimento ao cliente:** a consistência nos processos leva a uma maior previsibilidade nos níveis de resposta aos clientes;
- **Flexibilidade:** a utilização de *software* para controlar os processos permite um redesenho alinhado às necessidades de mudança no negócio;
- **Melhoria de processos de negócios:** o foco nos processos de negócio leva a sua racionalização e simplificação;

Alguns SGWs de código aberto disponíveis podem ser citados, tais como o WfMOpen, que é uma ferramenta de *workflow* baseada em java, o Taverna, criado para o projeto e execução de *workflows* científicos e o Kepler, que também é um conjunto de ferramentas para o projeto, execução e compartilhamento de *workflows* científicos. Apesar de existirem diversas ferramentas, é preferível o desenvolvimento de uma que esteja alinhada às necessidades da organização.

Sendo assim, para a implementação de um sistema de controle de *workflow*, é desejável a aplicação de uma metodologia ágil, que visa a satisfação do cliente mediante a entrega inicial e contínua de componentes úteis de software, desenvolvidos por um processo iterativo, utilizando o mínimo de recursos. É desejável, também, a utilização de uma plataforma de software livre baseada em *web* integrada a um banco de dados, com a finalidade de garantir adaptabilidade e acesso fácil aos dados. Diante dessas considerações, o *framework* voltado para *web Ruby on Rails* mostra-se uma ótima ferramenta para se construir um SGW.

4.3.1 Ruby on Rails

O *Ruby on Rails* é um *framework* livre de desenvolvimento *web* escrito na linguagem de programação Ruby [65]. Esse *framework* adota o padrão de arquitetura *Model-View-Controller* (MVC) ou Modelo, Visão e Controle. Os Modelos são responsáveis pela conexão com o banco de dados, geração de consultas e pelas regras de associação e validação dos dados. O Controle é a camada intermediária entre a *web* e o sistema, ou seja, possui as funções de traduzir e transmitir os dados para os modelos e de preparar os resultados para a camada de Visão. A Visão é a camada onde os resultados das operações e os dados são apresentados [66]. A Figura 17 representa o esquema da arquitetura MVC adotada pelo *Ruby on Rails*.

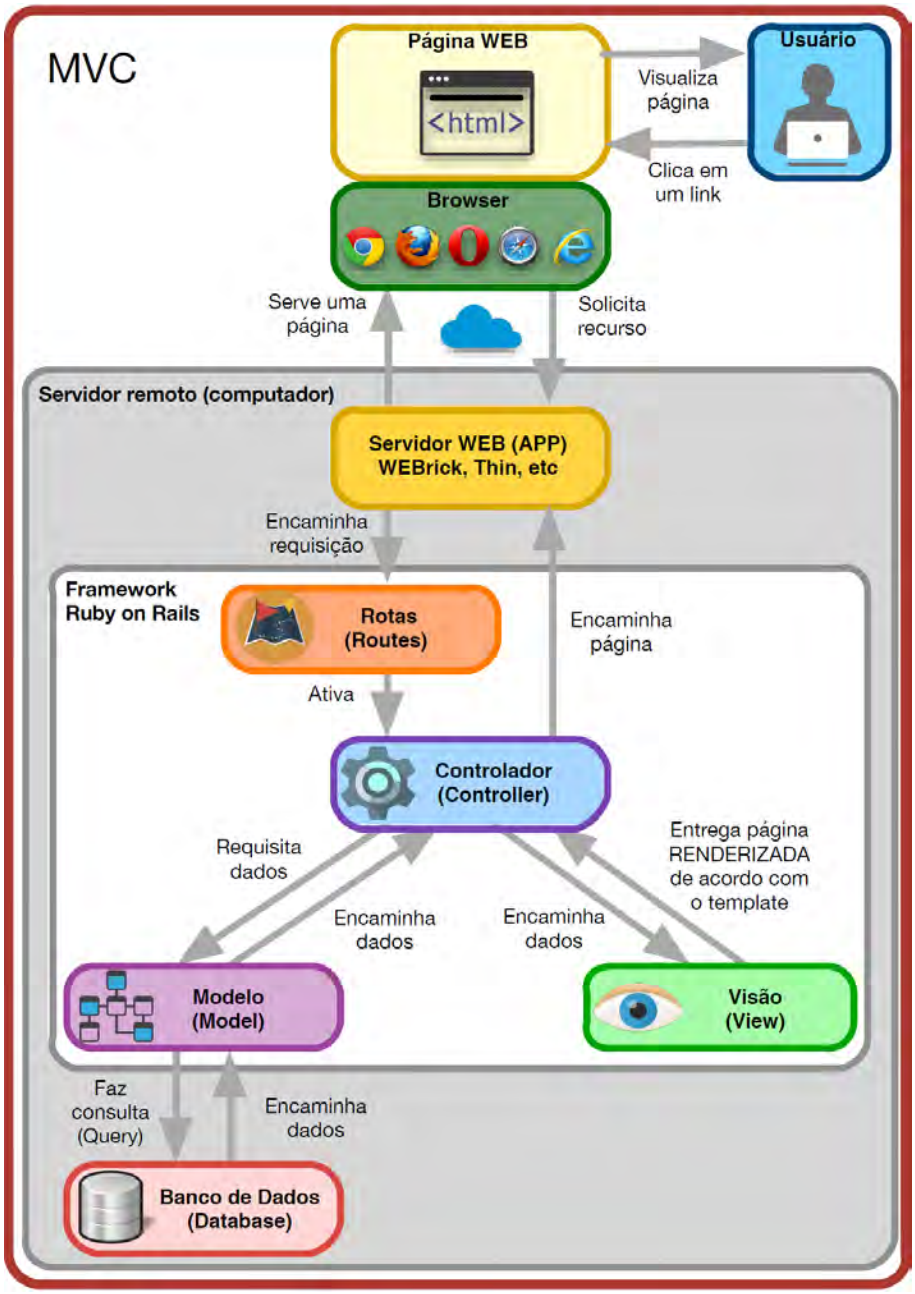
Ao interagir com um aplicativo Rails, o navegador envia uma requisição, que é recebida por um servidor *web* e repassada para um controlador Rails conforme indicado nas rotas cadastradas. Em alguns casos, o controlador cria uma visão de forma direta, retornando para o usuário. Porém, é mais comum o controlador interagir antes com um modelo e só depois disso processar a visão requisitada [65].

As vantagens em se utilizar o *Ruby on Rails* incluem a sua agilidade, a simplicidade da sintaxe da linguagem Ruby e a quantidade de métodos existentes, pois a maioria dos conceitos de desenvolvimento ágil está englobado nesse *framework*, o que facilita muito a programação. GitHub e Airbnb são exemplos de aplicações que foram construídas com *Ruby on Rails*.

Comparado a outros, o *Ruby on Rails* permite que as funcionalidades de um sistema possam ser implementadas de maneira incremental por conta de alguns padrões e conceitos adotados. Como pilares do Rails estão os conceitos de *Don't Repeat Yourself* (DRY) e *Convention over Configuration* (CoC). O conceito de DRY incentiva a fazer bom uso da reutilização de código, que é também uma das principais vantagens da orientação a objetos. O conceito de CoC traz o benefício de poder escrever muito menos código para implementar uma determinada funcionalidade em nossa aplicação, desde que respeitados alguns padrões de nome e localização de arquivos, nome de classes e métodos, entre outras regras simples e fáceis de serem memorizadas e seguidas [67].

Algumas convenções são adotadas para nomes de variáveis e métodos, que são sempre minúsculos e separados por '_' (traço inferior). Variáveis com nomes em maiúsculo são sempre constantes. Os nomes das tabelas seguem sempre o plural, os nomes dos modelos seguem sempre o singular e os nomes dos controles seguem sempre o plural. Por exemplo: criando o modelo `Order`, a tabela criada terá o nome de `orders` e o controle criado terá o nome de `OrdersController`.

Figura 17 - Representação esquemática do padrão de arquitetura MVC adotado pelo *Ruby on Rails*



Fonte: [66].

A conexão entre a aplicação e o banco de dados também é preparada pelo *Ruby on Rails*, que é compatível com uma grande quantidade de bancos, tais como MySQL e PostgreSQL. Por padrão, o Rails utiliza o SQLite3, pois é bem simples, versátil e consegue se comportar bem em um ambiente de desenvolvimento, além de possibilitar

que as informações sejam salvas em qualquer local do sistema de arquivos [67]. Para manipular os dados do banco de dados, pode-se utilizar o código Ruby através da console, portanto, não é preciso ter conhecimento da sintaxe SQL. A Tabela 5 apresenta quais são os principais tipos de dados suportados pelo Rails, que são passados na criação do modelo através do comando `rails generate model Nome campo1:tipo campo2:tipo ....`

Tabela 5 - Principais tipos de dados suportados pelo Rails

Tipo	**Significado**
:string	Texto curto
:text	Texto longo
:integer	Inteiro
:float	Fracionário (ponto flutuante)
:datetime	Data/Hora
:date	Data
:time	Hora
:binary	Para Armazenar Arquivos
:boolean	Verdadeiro ou Falso
:references	Para estabelecer associações

Fonte: [66].

Para validar esses dados, pode-se utilizar o método `validate` no formato `validates :campo, validação1, validação2, validação3 ...` nos Modelos criados. A Tabela 6 apresenta as validações mais comuns fornecidas pelo Rails.

Tabela 6 - Validações mais comuns fornecidas pelo Rails

Validação	**Significado**
presence: true	Não pode estar em branco
length: mininum:x, maximum:y	Precisa ter entre x e y caracteres
uniqueness: true	Precisa ser único na base de dados
numericality: true	Precisa ser numérico
acceptance: true	O usuário precisa ter marcado a checkbox correspondente no formulário da view
confirmation: true	O usuário precisa preencher o mesmo valor em dois campos do formulário. Muito usado para cadastro de senhas ou emails
format: with: Regex	Verifica formatação de um campo de acordo com uma expressão regular. Existem muitas expressões de validação prontas na internet
inclusion: in: x..y	Aceita valores entre x e y
inclusion: in: %w(A,B,C)	Aceita apenas os termos A, B ou C (usado para texto)
exclusion: in: ...	Não aceita os valores especificados. A forma de especificar os valores é análoga à validação inclusion
message: "mensagem"	Opção que pode ser incluída em qualquer das validações acima para substituir a mensagem de erro padrão exibida ao usuário quando a validação não é atendida

Fonte: [66].

Quando diversos modelos se relacionam na aplicação, é preciso que o tipo de relacionamento entre eles seja informado ao Rails. A Tabela 7 apresenta os relacionamentos que o Rails disponibiliza. As associações devem ser incluídas no modelo.

Tabela 7 - Associações fornecidas pelo Rails

Associação	**Significado**	**Exemplo**
belongs_to:model	Pertence à	Comentário publicação belongs_to:post
has_many:model	Tem muitos	Publicação tem muitos comentários has_many:comments
has_one:model	Tem um	Usuário tem uma mãe has_one:mother
has_many:model1, through: :model2	Possui muitos por intermédio de	Médico tem muitos pacientes por intermédio de consultas has_many :patients, through: :appointments
has_one:model1, through: :model2	Possui um por intermédio de	Fornecedor tem um histórico de conta por intermédio de conta has_one:account_history, through: :account
has_and_belongs_to_many:model	Possui e pertence à muitos	Cientista tem e pertence à muitos projetos has_and_belongs_to_many:projects

Fonte: [66].

Visando a facilitar a execução de tarefas repetitivas, o Rails possui um método designado *scaffold*, que facilita a construção das operações básicas em banco de dados *Create, Read, Update e Delete*, conhecido como CRUD, que em português significa Criar, Ler, Atualizar e Deletar. Esse método é muito útil na geração de um cadastro básico de alguma entidade em uma aplicação. Através do comando `rails generate scaffold nome campo1:tipo campo2:tipo ...`, é gerado todo o código necessário para listar todos os registros, exibir, criar, atualizar ou excluir um registro [67]. A Tabela 8 apresenta as ações padrão de um controle criadas automaticamente pelo Rails na geração de um *scaffold*.

Tabela 8 - Ações padrão de um controle criadas automaticamente na geração de um *scaffold*

Ação e Operação CRUD	**Objetivo**	**Receita CRUD (Ex.)**
index (R)	Renderizar a view que exibe todos os itens do modelo	@posts=Post.all
show (R)	Renderizar a view que exibe um item do modelo, cujo id é informado no endereço web	@post=Post.find(param[:id])
new (R)	Renderizar a view que exibe um formulário para criação de um elemento no modelo	@post=Post.new
edit (R)	Renderizar a view que exibe um formulário com os dados de um item do modelo para edição, cujo id é informado no endereço web	@post=Post.find(param[:id])
update (U)	Recebe os dados provenientes do formulário da view edit e atualiza o elemento correspondente no modelo	@post.update(post_params)
create (C)	Recebe os dados provenientes do formulário da view new e cria um elemento correspondente no modelo	@post=Post.new(post_params) @post.save
destroy (D)	Exclui um elemento do modelo, cujo id é informado no endereço web. A ação é chamada em um link em uma das views (no código gerado pelo Scaffold, na view index, por exemplo)	@post=Post.find(param[:id]) @post.destroy

Fonte: [68].

O próprio conjunto de ferramentas embutidas em uma instalação padrão de um ambiente Rails 5.0 já inclui um servidor simples, o Puma, suficiente para que o desenvolvedor possa acessar sua aplicação através do *browser*. Esse servidor não é indicado para aplicações em produção por ser muito simples e não contar com recursos mais avançados necessários para disponibilizar uma aplicação para um grande número de usuários [67].

A partir do Rails 3.1 foi criado um conceito para facilitar a extensão de uma aplicação chamado Rails *Engine*. Uma *engine* é uma espécie de mini aplicação Rails que permite isolar funcionalidades que podem ser reaproveitadas em qualquer aplicação Rails de uma forma estruturada e plugável, sendo conhecidas por gems. As *engines* devem ser adicionadas ao arquivo `Gemfile`, que é criado com a aplicação Rails e utilizado para descrever as gems e suas versões. Em seguida o comando `bundle install` deve ser executado para a instalação e atualização dessas gems, resolvendo facilmente as dependências e criando ambientes homogêneos em máquinas diferentes. Toda a funcionalidade das gems são fornecidas sem a necessidade de nenhum tipo de configuração adicional [67].

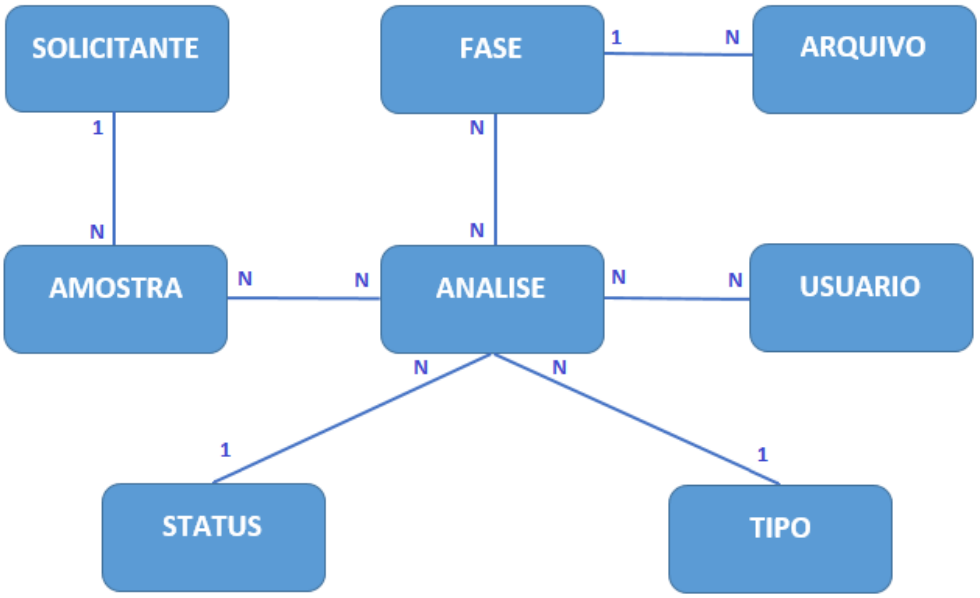
4.3.2 Proposta para um Sistema de Controle de *workflow* em Ruby on Rails

Considerando as vantagens que um SGW pode agregar, é proposta a implementação de um sistema que forneça controle sobre o processo de tipificação HLA do HLA-UERJ e que permita tanto a adição de novas funcionalidades de acordo com a necessidade do laboratório quanto a agregação de outros processos existentes.

Inicialmente foram identificadas algumas funcionalidades importantes que devem fazer parte do sistema proposto, tais como criar análises, incluir amostras, adicionar arquivos, cadastrar hemocentros e/ou médicos (solicitante da análise) e criar usuários.

A Figura 18 é uma representação gráfica simplificada do sistema proposto, no qual cada análise cadastrada deverá ser de um tipo, possuir um status e estar em uma determinada fase de acordo com o seu tipo. Cada análise criada deverá ainda, possuir um ou mais usuários e nenhum ou mais arquivos anexados. Deverá possuir uma ou mais amostras e cada amostra terá um solicitante (hemocentro ou médico) vinculado.

Figura 18 - Representação gráfica simplificada do sistema proposto



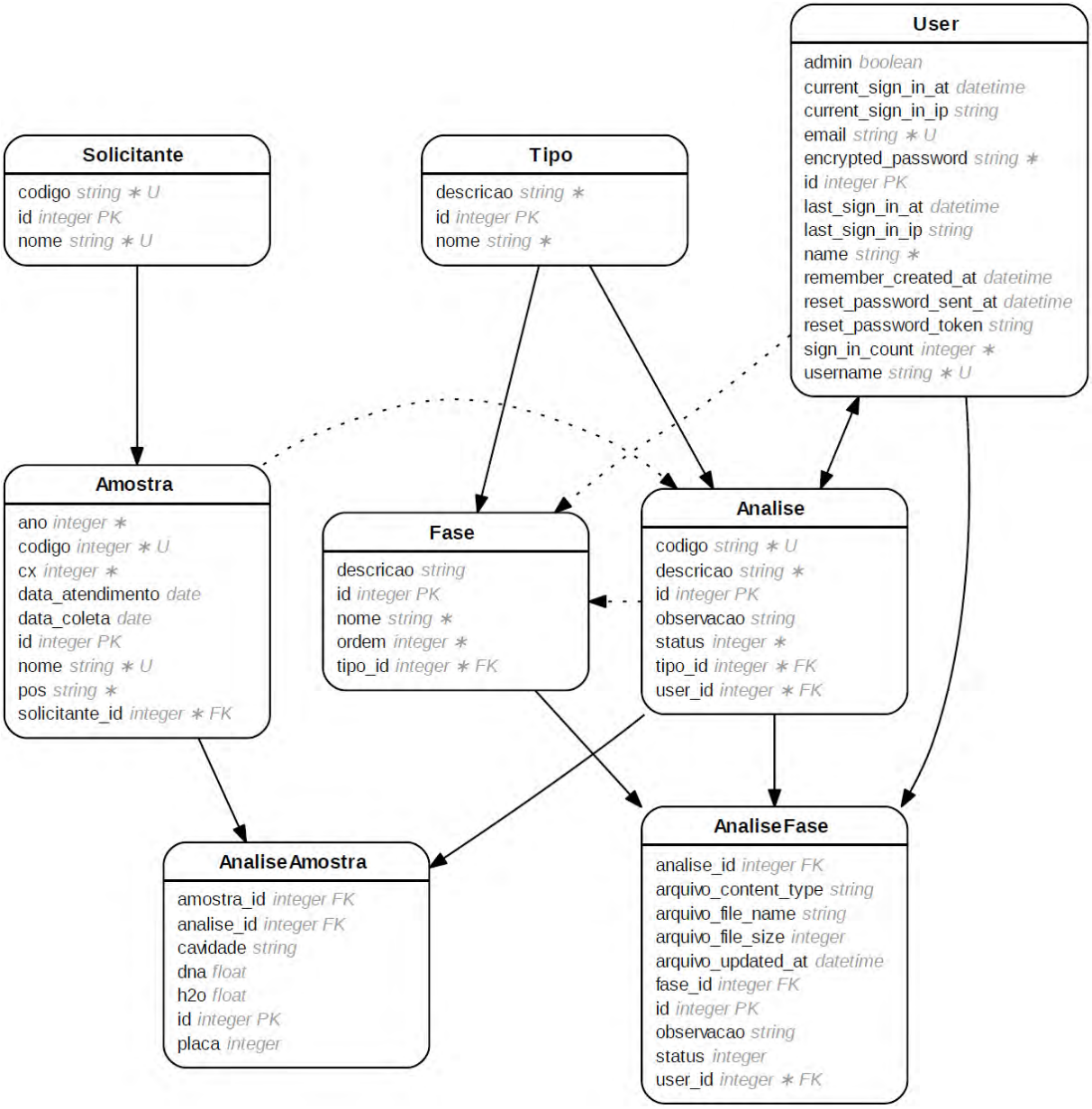
Fonte: O autor, 2017.

Os diagramas das tabelas apresentadas foram geradas utilizando a `gem 'rails-erd'`, nos quais os campos marcados com (*) são de preenchimento obrigatório, os marcados com (U) são campos de dados únicos, os marcados com (PK) são chaves primárias e os

marcados com (FK) são chaves estrangeiras.

A Figura 19 apresenta o diagrama esquemático dos modelos propostos gerado utilizando a `gem 'rails-erd'`. No diagrama as setas simples indicam associações do tipo um-para-muitos e as setas duplas indicam associações do tipo muitos-para-muitos. As associações muitos-para-muitos que utilizam uma tabela de junção são representadas por setas pontilhadas.

Figura 19 - Diagrama esquemático dos modelos propostos



Fonte: O autor, 2017.

Abaixo seguem as associações para os modelos propostos:

```
class User
    has_many :analises
    has_many :analise_fases
    has_many :fases, through: :analise_fases

class Tipo
    has_many :fases
    has_many :analises

class Fase
    belongs_to :tipo
    has_many :analise_fases
    has_many :analises, through: :analise_fases
    has_many :users, through: :analise_fases

class Analise
    belongs_to :user
    belongs_to :tipo
    has_many :analise_amostras
    has_many :amostras, through: :analise_amostras
    has_many :analise_fases
    has_many :fases, through: :analise_fases
    has_many :users, through: :analise_fases

class AnaliseFase
    belongs_to :fase
    belongs_to :user
    belongs_to :analise

class Solicitante
    has_many :amostras
```

```
class Amostra
    belongs_to :solicitante
    has_many :analise_amostras
    has_many :analises, through: :analise_amostras

class AnaliseAmostra
    belongs_to :analise
    belongs_to :amostra
```

A primeira funcionalidade que o sistema proposto deve possuir é a de multiusuários, o que permite a utilização do sistema por vários usuários. A princípio pensou-se em dividir os usuários em dois tipos: usuários comuns e usuários administradores. Somente os usuários administradores deverão ter as permissões de criar, editar e deletar usuários. Cada usuário deverá se autenticar no sistema para acessar as funcionalidades permitidas.

Sugere-se a geração de um `scaffold` para o modelo de usuários `User` com o campo `admin` sendo do tipo `boolean`, indicando se o usuário é ou não administrador, o campo `name` do tipo `string` e o campo `username` também do tipo `string` para identificação do usuário. O comando a ser utilizado seria: `rails g scaffold user admin:boolean name:string username:string`, o que já criaria a tabela `users` e os métodos para criar, exibir, editar e deletar usuários.

Para a autenticação dos usuários sugere-se a utilização da `gem 'devise'` juntamente com a `gem 'cancancan'` a fim de fornecer os devidos acessos aos usuários corretos. A `gem 'devise'` é instalada sobre o modelo `User` através do comando `rails generate devise user`, o que inclui os demais campos da tabela `users`, tais como `email` e `password`, que são os utilizados por padrão para autenticação. A `gem 'cancancan'` possibilita a configuração das permissões de acordo com o tipo de usuário. Quando instalada e gerada através do comando `rails generate cancan:ability`, é gerado o arquivo `ability.rb` na pasta de modelos e é nesse arquivo que as permissões devem ser configuradas utilizando o método `can`. Esse método possui dois parâmetros: um é a ação que desejamos realizar e o segundo é a classe de modelo em que a ação será aplicada. No caso do sistema, se o usuário for administrador, deverá ser capaz de gerenciar todos os usuários. Um exemplo da utilização seria:

```
if user.admin?
    can :manage, :all
end
```

Para utilização do método, também é necessário que seja incluído o código `include CanCan::ControllerAdditions` no arquivo de controle da aplicação, que é o arquivo `aplication_controller.rb`.

Os campos da tabela `users` proposta são apresentados na Figura 20 e descritos na Tabela 9. O campo `admin` indica se o usuário é administrador ou não, ou seja, se o valor do campo é `true`, o usuário é administrador, se é `false`, o usuário é do tipo comum.

Figura 20 - Tabela do modelo `User`

Fonte: O autor, 2017.

Tabela 9 - Descrição dos campos do modelo `User`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
admin	campo para identificação do tipo de usuário. Se o valor for verdadeiro o usuário é do tipo administrador. Se o valor for falso o usuário é do tipo comum.
username	campo para o apelido do usuário
name	campo para o nome completo do usuário
email	campos criados automaticamente pela gem devise para a manipulação dos usuários
encrypted_password	
current_sign_in_at	
current_sign_in_ip	
last_sign_in_at	
last_sign_in_ip	
remember_created_at	
reset_password_sent_at	
reset_password_token	
sign_in_count	

Fonte: O autor, 2017.

O campo `username` é de presença obrigatória e único e permite em sua composição tanto números quanto letras minúsculas e tamanho entre 4 e 10 caracteres. O campo `name` também é de presença obrigatória e permite em sua composição letras minúsculas e maiúsculas e tamanho entre 4 e 30 caracteres. Abaixo estão as validações para o modelo `User` proposto.

```
validates :username, :presence => true
validates :username, :uniqueness => true
validates :username, :format => /[0-9a-z]/
validates :username, length: { in: 4..10 }
validates :name, :presence => true
validates :name, :format => /[a-zA-Z]/
validates :name, length: { in: 4..30 }
```

Outra funcionalidade que o sistema deverá possuir é a de cadastrar os tipos de análise e suas respectivas fases. Para tal, deverá ser gerado um `scaffold` com os campos

`nome` do tipo `string` e o campo `descrição` do tipo `string`. A Figura 21 apresenta e a Tabela 10 descreve os campos da tabela `tipos` proposta.

Figura 21 - Tabela do modelo `Tipo`

Fonte: O autor, 2017.

Tabela 10 - Descrição dos campos do modelo `Tipo`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
nome	campo para o nome do tipo de análise
descricao	campo para descrição do tipo de análise

Fonte: O autor, 2017.

O campo `nome` é de presença obrigatória e permite em sua composição letras minúsculas e maiúsculas e tamanho entre 4 e 30 caracteres. O campo `descricao` é de presença obrigatória e permite tamanho entre 1 e 50 caracteres. Abaixo estão as validações para o modelo `Tipo` proposto.

```
validates :nome, :presence => true
validates :nome, :format => /[a-zA-Z]/
validates :nome, length: { in: 4..30 }
validates :descricao, :presence => true
validates :descricao, length: { in: 1..50 }
```

Para cada tipo de análise criado, uma ou demais fases deverão ser criadas. Para tal, deverá ser gerado um `scaffold` com os campos `ordem` do tipo `integer`, o campo `nome` do tipo `string`, o campo `descrição` do tipo `string` e o campo `tipo` como referência

ao modelo `Tipo`. O campo `ordem` existe para que seja estabelecida uma sequência de execução das fases cadastradas. A Figura 22 apresenta e a Tabela 11 descreve os campos da tabela `fases` proposta.

Figura 22 - Tabela do modelo `Fase`

Fonte: O autor, 2017.

Tabela 11 - Descrição dos campos do modelo `Fase`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
ordem	campo para o número de sequência da fase
nome	campo para o nome da fase
descricao	campo para descrição da fase
tipo_id	campo de referência para o modelo Tipo

Fonte: O autor, 2017.

O campo `nome` é de presença obrigatória e permite em sua composição letras minúsculas e maiúsculas e tamanho entre 4 e 50 caracteres. O campo `descricao` permite tamanho máximo de 50 caracteres. Os campos `ordem` e o campo de referência `tipo_id` são de presença obrigatória. Abaixo estão as validações para o modelo `Fase` proposto.

```
validates :nome, :presence => true
validates :nome, :format => /[a-zA-Z]/
validates :nome, length: { in: 4..50 }
validates :descricao, length: { maximum: 50 }
validates :ordem, :presence => true
validates :tipo_id, :presence => true
```

De acordo com o mapeamento realizado das atividades de tipificação HLA do HLA-UERJ, sugere-se que, inicialmente, seja criado o tipo "Tipagem HLA - Média Resolução" com as fases "Cadastro das amostras", "Extração de DNA", "Criação de mapa de identificação PCR", "Amplificação", "Verificação do padrão das amostras", "Atualização do mapa de identificação", "Hibridização", "Inclusão dos dados nos sistemas"e "Emissão de laudo" e o tipo "Tipagem HLA - Alta Resolução" com as fases "Cadastro das amostras", "Extração de DNA", "Criação de mapa de identificação PCR", "Amplificação", "Verificação do padrão das amostras", "Atualização do mapa de identificação", "Quantificação de DNA", "Ajuste de concentração das amostras", "Criação do mapa de identificação NGS", "Sequenciamento", "Inclusão dos dados nos sistemas" e "Emissão de laudo".

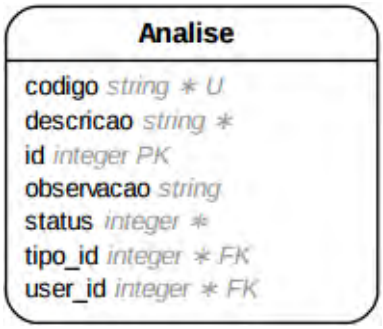
O sistema deverá possuir também, a funcionalidade de cadastro de análises. Cada análise deverá ser de um tipo, possuir um status, possuir vinculadas as fases que compõe o tipo de análise, a identificação do usuário criador da análise e um campo para observações. Para tal, deverá ser gerado um `scaffold` com os campos `codigo` do tipo `string`, o campo `descrição` do tipo `string`, o campo `observação` também do tipo `string`, o campo `status` do tipo `integer`, o campo `user` como referência ao modelo `User` e o campo `tipo` como referência ao modelo `Tipo`.

Para o campo `status` sugerem-se os seguintes valores: "criada", "em andamento", "concluída", "suspensa" e "cancelada". Esse campo pode ser criado utilizando o método `enum`, que é um tipo especial de dados que habilita uma variável ser configurada com constantes pré-definidas. Segue um exemplo de declaração que deve ser feita no arquivo de modelo `analise.rb`:

```
enum status:
[:criada,:'em andamento',:concluída,:suspensa,:cancelada]
```

A Figura 23 apresenta e a Tabela 12 descreve os campos da tabela `analises`.

Figura 23 - Tabela do modelo `Analise`



Fonte: O autor, 2017.

Tabela 12 - Descrição dos campos do modelo `Analise`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
codigo	campo de código da análise para controle interno (independente da chave primária)
descricao	campo para descrição da análise
status	campo para identificação do estado da análise
observacao	campo para inclusão de observações sobre a análise
tipo_id	campo de referência para o modelo Tipo
user_id	campo de referência para o modelo User

Fonte: O autor, 2017.

O campo `codigo` é de presença obrigatória e único e permite em sua composição tanto números quanto letras maiúsculas e minúsculas e tamanho entre 1 e 10 caracteres. O campo `descricao` também é de presença obrigatória e permite tamanho entre 1 e 50 caracteres. O campo `observacao` permite tamanho entre 0 e 100 caracteres. O campo `status` e os campos de referência `user_id` e `tipo_id` são de presença obrigatória. Abaixo estão as validações para o modelo `Analise` proposto.

```
validates :codigo, :presence => true
validates :codigo, :uniqueness => true
validates :codigo, :format => /[0-9a-zA-z]/
```

```
validates :codigo, length: { in: 1..10 }
validates :descricao, :presence => true
validates :descricao, length: { in: 1..50 }
validates :observacao, length: { in: 0..100 }
validates :status, :presence => true
validates :user_id, :presence => true
validates :tipo_id, :presence => true
```

As fases da análise deverão ser vinculadas automaticamente na criação da análise, de acordo com o tipo selecionado. Cada fase vinculada à análise deverá possuir um status, um usuário responsável, que não necessariamente precisa ser o mesmo usuário que criou a análise no sistema e um campo para observações. Além disso, para cada fase, o sistema deverá permitir a inclusão de arquivos, que poderão ser desde arquivos de texto à arquivos de foto documentação, incluindo os arquivos de mapa de localização e identificação das amostras. Os arquivos deverão estar disponíveis tanto para visualização quanto para *download*.

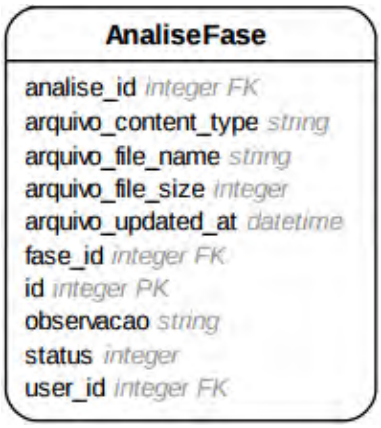
Para tal, deverá ser gerada uma *join table*, que também é chamada de tabela de junção, entre os modelos `Analise` e `Fase`. O nome da tabela é a concatenação do nome dos dois modelos, que no caso será `analise_fases`. O modelo `AnaliseFase` deverá possuir o campo `analise` como referência ao modelo `Analise`, o campo `fase` como referência ao modelo `Fase`, o campo `user` como referência ao modelo `User`, o campo `observacao` do tipo `string`, o campo `status` do tipo `integer` e o campo `arquivo` do tipo `attachment`. Para a geração do campo `arquivo`, sugere-se a utilização da `gem 'paperclip'`. No arquivo de modelo `analise_fase.rb` deverá ser incluída a linha `has_attached_file :arquivo`.

Para o campo `status` sugerem-se os seguintes valores: "criada", "em andamento" e "concluída". Esse campo pode ser criado utilizando o método `enum`. Segue um exemplo de declaração que deve ser feita no arquivo de modelo `analise_fase.rb`:

```
enum status:
[:criada,:'em andamento',:concluída]
```

A Figura 24 apresenta e a Tabela 13 descreve os campos da tabela `analise_fases` proposta.

Figura 24 - Tabela do modelo `AnaliseFase`



Fonte: O autor, 2017.

Tabela 13 - Descrição dos campos do modelo `AnaliseFase`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
analise_id	campo de referência para o modelo Analise
fase_id	campo de referência para o modelo Fase
user_id	campo de referência para o modelo User
observacao	campo para inclusão de observações sobre a fase da análise
status	campo para identificação do estado da fase da análise
arquivo_content_type	campos criados automaticamente pela gem paperclip para a manipulação de arquivos
arquivo_file_name	
arquivo_file_size	
arquivo_updated_at	

Fonte: O autor, 2017.

O campo `observacao` permite tamanho entre 0 e 100 caracteres e os campos de referência `user_id`, `analise_id` e `fase_id` são de presença obrigatória. Abaixo estão as validações para o modelo `AnaliseFase` proposto.

```
validates :observacao, length: { in: 0..100 }
validates :user_id, :presence => true
validates :analise_id, :presence => true
validates :fase_id, :presence => true
```

A Figura 25 apresenta uma proposta de tela para alteração do fluxo das análises.

Figura 25 - Proposta de tela para alteração do fluxo das análises

HLA-UERJ - Editar Análise

Olá, usuario1 | Sair | Trocar senha | Administrar Usuários | Home

Código
A123

Tipo:
Tipagem HLA - Média Resolução

Descrição
Analise I

Estado
em andamento

Observação

	Fase	Usuário Resp.	Estado	Observação	Arquivos		
1	Cadastro das amostras	usuario1	concluída	Amostras cadastradas	Lista_de_Amostras.xlsx	Escolher arquivo	Cadastrar Amostras
2	Extração de DNA	usuario2	concluída	ok	Saida_Centrifuga.txt	Escolher arquivo	
3	Criação de mapa de identificação PCR	usuario1	em andamento			Escolher arquivo	Gerar mapa
4	Amplificação	usuario1	criada			Escolher arquivo	
5	Verificação do padrão das amostras	usuario2	criada			Escolher arquivo	
6	Atualização do mapa de identificação	usuario2	criada			Escolher arquivo	Atualizar mapa
7	Hibridização	usuario1	criada			Escolher arquivo	
8	Inclusão dos dados nos sistemas	usuario3	criada			Escolher arquivo	
9	Emissão de Laudo	usuario3	criada			Escolher arquivo	

Salvar Análise

Exibir | Voltar

Fonte: O autor, 2017.

Outra proposta mais dinâmica de alteração do fluxo das análises seria apresentar na tela de forma ativa apenas a fase atual com um botão que permitisse a sua conclusão e consequentemente alterasse o estado da fase seguinte de "criada" para "em andamento", passando esta última a ser a nova fase atual automaticamente.

Cada análise deverá permitir a vinculação de uma ou mais amostras, sejam elas novas ou já anteriormente cadastradas. Cada amostra vinculada a análise, deverá pertencer a um solicitante, que pode ser tanto um hemocentro quanto um médico. Para a funcionalidade de cadastrar solicitante, deverá ser gerado um `scaffold` com os campos `codigo` do tipo `string`, para identificação do hemocentro ou médico e o campo `nome` do tipo `string`. A Figura 26 apresenta e a Tabela 14 descreve os campos da tabela `solicitantes` proposta.

Figura 26 - Tabela do modelo `Solicitante`

Fonte: O autor, 2017.

Tabela 14 - Descrição dos campos do modelo `Solicitante`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
codigo	campo de código do solicitante para controle interno (independente da chave primária)
nome	campo para o nome do solicitante

Fonte: O autor, 2017.

O campo `codigo` é de presença obrigatória, único e tamanho entre 4 e 10 caracteres. O campo `nome` também é de presença obrigatória, único e permite em sua composição letras maiúsculas e minúsculas e tamanho entre 4 e 30 caracteres. Abaixo estão as validações para o modelo `Solicitante` proposto.

```
validates :codigo, :presence => true
validates :codigo, :uniqueness => true
validates :codigo, length: { in: 4..10 }
validates :nome, :presence => true
validates :nome, :uniqueness => true
validates :nome, :format => /[a-zA-Z]/
validates :nome, length: { in: 4..30 }
```

Para o cadastro de amostras um `scaffold` deverá ser gerado com os campos `codigo` do tipo `integer`, o campo `cx` do tipo `integer`, o campo `ano` do tipo `integer`, o campo `pos` do tipo `string`, o campo `nome` do tipo `string`, o campo `data_coleta` do tipo `date`,

o campo `data_atendimento` do tipo `date`, o campo `solicitante_id` como referencia ao modelo `Solicitante`. A Figura 27 apresenta e a Tabela 15 descreve os campos da tabela `amostras` proposta.

Figura 27 - Tabela do modelo `Amostra`

Fonte: O autor, 2017.

Tabela 15 - Descrição dos campos do modelo `Amostra`

Campo	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
codigo	campo para o código da amostra recebida (independente da chave primária)
nome	campo para o nome do paciente ou doador da amostra recebida
data_atendimento	campo para a data de atendimento
data_coleta	campo para a data da coleta
cx	campos para identificação da amostra recebida (caixa, ano e posição)
ano	
pos	
solicitante_id	campo de referência para o modelo Solicitante

Fonte: O autor, 2017.

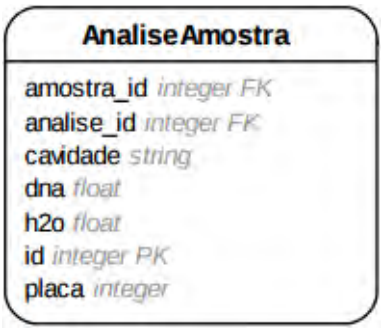
O campo `codigo` é de presença obrigatória, único e tamanho entre 4 e 10 caracteres. O campo `nome` é de presença obrigatória e permite em sua composição letras minúsculas e maiúsculas e tamanho entre 4 e 30 caracteres. O campo `cx` é de presença obrigatória e tamanho entre 1 e 3 caracteres. O campo `ano` é de presença obrigatória e tamanho de 2

caracteres. O campo `pos` é de presença obrigatória permite em sua composição números e letras maiúsculas e minúsculas e tamanho entre 2 e 3 caracteres. O campo de referência `solicitante_id` é de presença obrigatória. Abaixo estão as validações para o modelo `Amostra` proposto.

```
validates :codigo, :presence => true
validates :codigo, :uniqueness => true
validates :codigo, length: { in: 4..10 }
validates :cx, :presence => true,
validates :cx, length: { in: 1..3 }
validates :ano, :presence => true
validates :ano, length: { is: 2 }
validates :pos, :presence => true
validates :pos, length: { in: 2..3 }
validates :pos, :format => /[0-9a-hA-H]/
validates :nome, :presence => true
validates :nome, :format => /[a-zA-Z]/
validates :nome, length: { in: 4..30 }
validates :solicitante_id, :presence => true
```

Para a vinculação das amostras à análise, deverá ser gerada uma tabela de junção entre os modelos `Analise` e `Amostra` de nome `analise_amostras`. O modelo `AnaliseAmostra` deverá possuir o campo `analise` como referência ao modelo `Analise`, o campo `amostra` como referência ao modelo `Amostra`, o campo `dna` do tipo `float`, o campo `h2o` do tipo `float`, o campo `cavidade` do tipo `string` e o campo `placa` do tipo `integer`. Os campos `dna` e `h2o` deverão ser preenchidos nas fases de "Quantificação de DNA" e de "Ajuste de concentração das amostras", respectivamente. A Figura 28 apresenta e a Tabela 16 descreve os campos da tabela `analise_amostras` proposta.

Figura 28 - Tabela do modelo `AnaliseAmostra`



Fonte: O autor, 2017.

Tabela 16 - Descrição dos campos do modelo `AnaliseAmostra`

Campos	**Descrição**
id	campo de chave primária de auto incremento gerado automaticamente pelo Rails na criação do modelo
amostra_id	campo de referência para o modelo Amostra
analise_id	campo de referência para o modelo Analise
placa	campo para identificação da microplaca
cavidade	campo para identificação da cavidade da microplaca
dna	campo para a quantificação de DNA da amostra
h2o	campo para realizar o ajuste de concentração da amostra

Fonte: O autor, 2017.

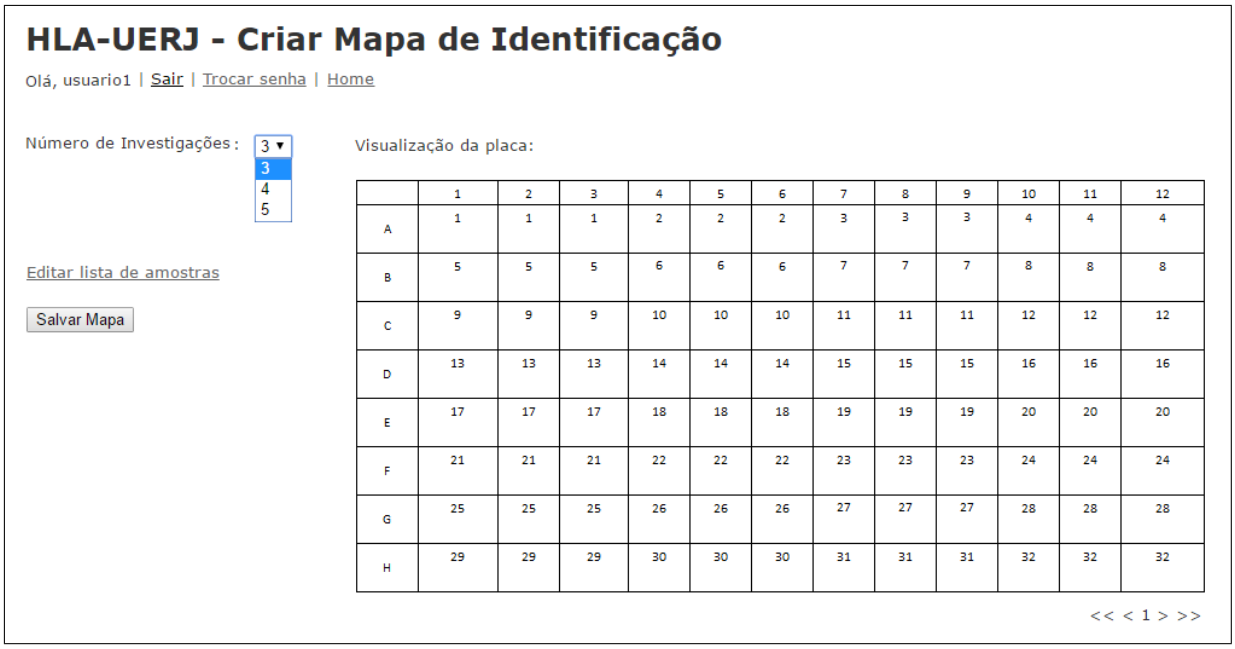
O campo `cavidade` permite em sua composição números e letras maiúsculas e minúsculas e tamanho máximo de 3 caracteres. Os campos de referência `analise_id` e `amostra_id` são de presença obrigatória. Abaixo estão as validações para o modelo `AnaliseAmostra` proposto.

```
validates :cavidade, length: { in: 0..3 }
validates :cavidade, :format => /[0-9a-hA-H]/
validates :analise_id, :presence => true
validates :amostra_id, :presence => true
```

Na fase de "Criação de mapa de identificação" o sistema deverá ser capaz de gerar e armazenar o mapa de localização e identificação das amostras na microplaca de forma

automática. A funcionalidade deverá considerar a quantidade de investigações a serem realizadas. Através de uma tela de seleção das amostras e de acordo com uma caixa de seleção indicando o número de investigações desejadas, a localização das amostras já deverá ser gerada de forma automática e um desenho do mapa deverá ser visualizado na tela do sistema. A fim de garantir a mobilidade do mapa, um *tablet* poderá ser utilizado dentro das salas de análise do laboratório, evitando-se assim, o uso do papel impresso. Figura 29 apresenta uma proposta de tela para a criação do mapa de identificação e localização das amostras.

Figura 29 - Proposta de tela para a criação do mapa de identificação e localização das amostras



	1	2	3	4	5	6	7	8	9	10	11	12
A	1	1	1	2	2	2	3	3	3	4	4	4
B	5	5	5	6	6	6	7	7	7	8	8	8
C	9	9	9	10	10	10	11	11	11	12	12	12
D	13	13	13	14	14	14	15	15	15	16	16	16
E	17	17	17	18	18	18	19	19	19	20	20	20
F	21	21	21	22	22	22	23	23	23	24	24	24
G	25	25	25	26	26	26	27	27	27	28	28	28
H	29	29	29	30	30	30	31	31	31	32	32	32

Fonte: O autor, 2017.

Com o objetivo de exemplificar e iniciar o projeto do sistema proposto, foi desenvolvido um protótipo com as funcionalidades propostas de autenticação de multiusuários, de cadastro de tipos e fases e de cadastro de análises. Para o desenvolvimento do protótipo foi utilizado o sistema operacional Linux distribuição Ubuntu 14.04, a versão 5.0.1 do Rails, a versão 2.0 do Ruby, o banco de dados SQLite3, que é o banco padrão do Rails e as gems paperclip-5.0.0, cancancan-1.16.0, devise-4.2.0 e load_and_authorize_resource-0.4.0. As telas do protótipo são apresentadas no Apêndice - Telas do protótipo.

4.3.3 Proposta de arquitetura de alta disponibilidade

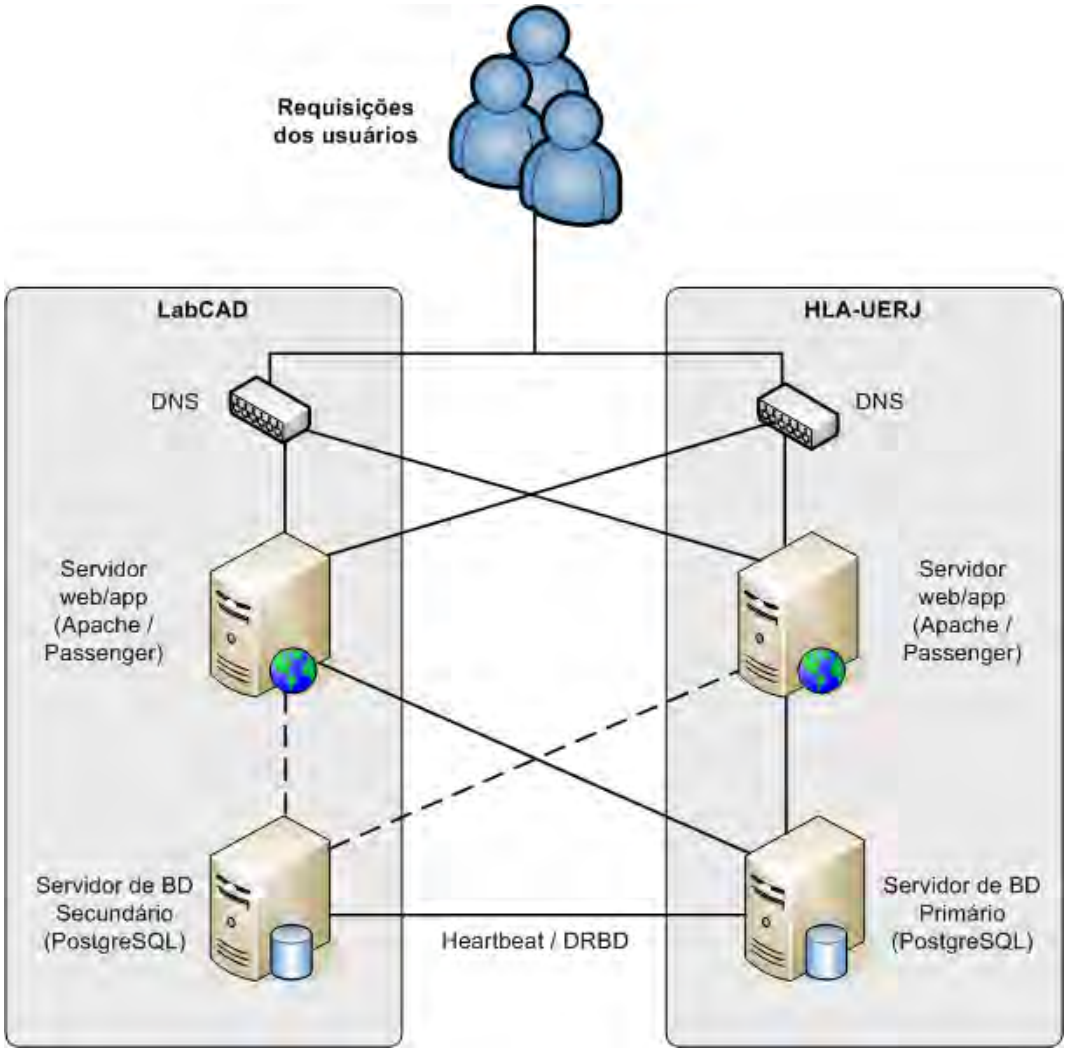
Com a finalidade de aumentar a confiabilidade do sistema de controle de *workflow*, é proposto que o mesmo seja implantado em um ambiente de alta disponibilidade. Um sistema altamente disponível é uma combinação de componentes de *hardware* e *software* que fornecem conjuntamente um serviço cuja disponibilidade é muito maior do que a disponibilidade de qualquer componente fornecendo serviço de forma individual [69].

Em ambientes computacionais falhas são inevitáveis e essas falhas podem ser tanto em razão de *hardware* quanto de *software*. Para suprir falhas em termos de *hardware* procura-se utilizar redundância de equipamentos, provendo duas ou mais cópias físicas do *hardware*. No que diz respeito a *software*, a falha em um nó é compensada ao migrar os recursos comprometidos para um outro nó em operação. Nesse contexto, a utilização de soluções de softwares livres pode proporcionar a redução dos custos totais.

A arquitetura proposta sugere que sejam utilizados tanto o ambiente do HLA-UERJ quanto o ambiente do LabCAD. O LabCAD está localizado no Pavilhão Reitor João Lyra Filho da UERJ e tem o intuito de abrigar pesquisadores da área de computação para fazer pesquisa de ferramental e técnicas de computação de alto desempenho, redes, arquitetura de computadores, linguagem, compiladores, entre outros, fornecendo infraestrutura para ensino de graduação e pós-graduação na área.

A Figura 30 apresenta um esquema gráfico da arquitetura de alta disponibilidade proposta. As linhas pontilhadas representam caminhos alternativos entre os servidores.

Figura 30 - Esquema gráfico da arquitetura de alta disponibilidade proposta



Fonte: O autor, 2017.

Para a comunicação entre os dois ambientes, deverá existir um *link* dedicado por meio de fibra óptica. Em cada ambiente deverá existir um servidor de DNS (*Domain Name System* ou Sistema de Nomes de Domínios, em português). O DNS é um serviço de resolução de nomes, sendo responsável por localizar e traduzir para números IPs nomes utilizados para designar serviços ou *hosts* de rede, tais quais os endereços dos sites que digitamos nos navegadores. Ambos deverão direcionar os acessos para os dois ambientes. Em caso de falha de um servidor, os DNSs já estarão apontados para os demais servidores da arquitetura.

Deverá existir em cada ambiente um servidor *web* responsável pela camada de apresentação. O *software* Apache poderá ser utilizado para a configuração dos servidores

web, pois além de ser robusto e de alta performance, é de fácil instalação e gratuito [70]. Cada servidor *web* acessará um servidor de aplicação e em caso de indisponibilidade do servidor primário, o servidor *web* redirecionará o acesso ao outro servidor de aplicação.

Buscando-se uma solução com baixo custo e como a quantidade de usuários é reduzida, sugere-se que tanto a camada de apresentação quanto a camada de aplicação sejam hospedadas na mesma máquina, porém é preferível que o banco de dados seja hospedado em máquina separada, já que em caso de falha, o banco de dados continua acessível pelos demais. A segregação do banco de dados também garante desempenho e segurança, já que o tipo de processamento realizado pelo banco diverge do realizado pelo servidor de apresentação e aplicação. O primeiro realiza mais operações de entrada e saída em disco ou memória e o segundo é mais voltado para criação de processos e comunicação de rede. Com relação à segurança, o fato de o banco de dados estar segregado, possibilita que o mesmo esteja em uma rede onde somente os servidores de apresentação e aplicação possam acessar.

Nesse contexto, ambos os servidores com a camada de apresentação e aplicação acessarão um único gerenciador de banco de dados, que poderá estar localizado em qualquer um dos dois ambientes. Esse, por sua vez, deverá ser replicado para o outro ambiente, seguindo o formato primário e secundário. Sugere-se que o servidor primário esteja no ambiente do HLA-UERJ, pois a maioria das requisições será realizada dentro desse ambiente.

Para os servidores, sugere-se a utilização do sistema operacional Linux, por se tratar de *software* livre. Para os servidores de aplicação, sugere-se que o *software* Phusion Passenger seja o utilizado, pois além de ter código livre, é um servidor de aplicação desenhado para ser rápido, robusto, leve, simples de ser configurado e compatível com Ruby. O Passenger é capaz de gerenciar processos e recursos, permitir a administração, monitoramento e diagnóstico de problemas e é compatível com Apache [71].

O banco de dados utilizado pode ser o PostgreSQL, que, apesar de ser de uso livre, é tão robusto e estável quanto as soluções comerciais [72]. Para garantir o sincronismo de dados entre o nó primário e o nó secundário, sugere-se a utilização da solução DRBD (*Distributed Replicated Block Device*). Não é sugerida, nessa proposta, a utilização de *storage* compartilhado entre os nós, pois seria um equipamento adicional que representaria um aumento significativo do custo de implementação.

O DRBD é uma solução de código aberto de replicação de blocos de disco (*Block-Device*) para Linux, na qual todas as alterações em uma sistema de arquivo são espelhadas em um sistema de arquivos residente em outro servidor. Age de forma transparente para aplicações e bancos de dados e evita a perda de dados durante falhas de hardware, possibilitando a recuperação automática do sistema sem a necessidade de outro equipamento adicional [73].

Para verificar a disponibilidade dos servidores pode ser utilizado um serviço de troca de mensagens. É sugerida a utilização do Heartbeat. O Heartbeat deve ser instalado tanto no servidor primário quanto no secundário. Assim, o servidor secundário envia mensagens ao primário e espera as respostas. Quando o primário não responde às mensagens, o processo de *failover*, que é o processo de migrar os serviços para um outro servidor, é iniciado de maneira automática pelo Heartbeat e o servidor secundário assume todas as funcionalidades do primário, inclusive o endereço IP.

Quando o antigo servidor primário volta a ficar disponível, é preciso que o mesmo seja informado que ele não é mais o primário. Esse mecanismo é conhecido como STONITH (*Shoot The Other Node In The Head*), que é necessário para evitar situações em que ambos os sistemas pensam que eles são os principais, o que levará à perda de dados [72].

Após a recuperação de uma falha, deve-se iniciar o processo de *failback*, que consiste em devolver ao servidor primário os serviços migrados durante a falha para o servidor secundário. Esse processo tem início com o DRBD realizando a sincronização dos discos e quando o servidor que teve a indisponibilidade voltar a estar disponível, as funcionalidades são devolvidas do secundário para o primário.

Como os dados representam informação valiosa, os bancos de dados do PostgreSQL devem sofrer *Backup* regularmente. Um dos métodos utilizados para rápida recuperação do banco é o SQL Dump. A idéia por trás do método de *dump* é gerar um arquivo de texto com comandos SQL que, quando retornados ao servidor, recriarão o banco de dados no mesmo estado em que estava no momento em que foi realizado o *backup* [72].

Para finalizar a proposta da arquitetura, é preciso destacar a importância do uso de sistemas *no-break* para suprir possíveis interrupções no fornecimento de energia elétrica.

CONCLUSÕES E TRABALHOS FUTUROS

Como foi colocado, a medicina laboratorial teve grande evolução nos últimos anos, possuindo papel fundamental na cadeia da saúde, sendo cada vez mais fonte geradora de informação com impacto na decisão clínica. Porém, tem enfrentado problemas com o armazenamento e recuperação dos dados atualmente.

A tecnologia da informação apresenta-se como uma importante ferramenta para solucionar tais questões e só traz benefícios. Dentre as vantagens destacam-se maior agilidade, segurança e qualidade técnica para as atividades, menor tempo de processamento de lotes e de correção do processo produtivo, menor índice de retrabalhos e erros, o que consequentemente, reduz o tempo de entrega. Entretanto, a cultura do papel representa um enorme obstáculo na aplicação da tecnologia da informação nos processos.

Dentro dessa perspectiva, o presente trabalho teve como objetivo inicial realizar o mapeamento das atividades de tipificação HLA exercidas pelo HLA-UERJ, o que permitiu adquirir uma visão ampla e precisa do processo como um todo. Com o processo de tipificação HLA detalhado e mapeado, foi possível observar e analisar a ocorrência de alguns riscos e vulnerabilidades no processo, tais como o controle das atividades ser descentralizado, o que prejudica uma rápida recuperação dos dados, dificulta o controle da redundância e põe em dúvida a integridade dos dados. O fato de o mapa de localização e identificação das microplacas ser preenchido à mão também implica na ocupação de espaço físico e torna a consulta difícil, demorada e inadequada.

Com esses elementos elencados, foi elaborada uma proposta de implementação de um sistema de controle de *workflow*, a fim de proporcionar ao processo maior acurácia, precisão, acessibilidade, rastreabilidade e confiabilidade dos dados, além de melhorar a comunicação e colaboração da equipe.

Para a proposta do sistema sugeriu-se a utilização do *framework Ruby on Rails* devido à sua simplicidade de sintaxe e a quantidade de métodos que já existem, além de ser baseado em *web* e ser livre. A plataforma mostrou-se bastante intuitiva, o que facilitou a modelagem do sistema proposto. Para exemplificar e incentivar o início da implementação, foi desenvolvido um protótipo com as funcionalidades de multiusuário, de gerenciamento de tipos de análises e respectivas fases e cadastro de análises sugeridas

na proposta.

Para a implantação do sistema, foi proposta uma arquitetura em alta disponibilidade, com custo reduzido, envolvendo tanto o ambiente do HLA-UERJ quanto o do LabCAD, buscando a utilização de poucos equipamentos e *softwares* de código livre.

Como trabalhos futuros, a expectativa é que os apontamentos e propostas sugeridas nesse trabalho possam servir como base para outros projetos, podendo o sistema proposto ser implementado na sua totalidade de modo que seja possível analisar a sua eficácia e o seu poder de otimização das atividades. Implementando-se o sistema, a arquitetura proposta poderá ser testada e avaliada. Outro ponto que pode ser considerado futuramente é a criação de novas funcionalidades para o sistema proposto, levando-se em consideração outros tipos de observações e análises dos demais processos do HLA-UERJ, já que o *Ruby on Rails* é totalmente escalável.

Para as novas funcionalidades, pode ser sugerido que seja incluído um controle de custos baseado no cadastro dos equipamentos e insumos utilizados em cada análise. Outra questão importante, seria a inclusão de um controle de *logs*, ou seja, um controle de todos os eventos ocorridos no sistema, o que envolve o registro de acessos e alterações. Essa funcionalidade é essencial para trabalhos de auditoria.

Outra funcionalidade que pode ser incluída no sistema proposto é a de troca de mensagens entre os usuários com a finalidade de melhorar ainda mais a comunicação na equipe. O desenvolvimento de filtros de busca dos registros também será uma funcionalidade bastante útil quando o banco de dados alcançar dimensões maiores.

REFERÊNCIAS

[1] SETUBAL, J. C.; MEIDANIS, J. *Introduction to computational molecular biology.* [S.l.]: PWS Pub., 1997.

[2] SELZER, P. M.; MARHÖFER, R.; ROHWER, A. *Applied bioinformatics: an introduction.* [S.l.]: Springer Science & Business Media, 2008.

[3] NCBI. Genbank and wgs statistics. 2016. Disponível em: <http://www.ncbi.nlm.nih.gov/genbank/statistics/>. Acesso em: 11 ago. 2016.

[4] XIONG, J. *Essential bioinformatics.* [S.l.]: Cambridge University Press, 2006.

[5] GIBAS, C.; JAMBECK, P. *Developing bioinformatics computer skills.* [S.l.]: O'Reilly Media, Inc., 2001.

[6] HLA - Laboratório de Histocompatibilidade e Criopreservação - UERJ. Disponível em: <www.hla.uerj.br/>. Acesso em: 29 set. 2016.

[7] HARRINGTON, J. *Aperfeiçoando processos empresariais.* [S.l.]: Makron Books, 1993.

[8] VILLELA, C. S. S. Dissertação de Mestrado em Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, *Mapeamento de processos como ferramenta de reestruturação e aprendizagem organizacional.* 2000. Disponível em: <https://repositorio.ufsc.br/handle/123456789/78638>. Acesso em: 05 nov. 2016.

[9] DATZ, D.; MELO, A. C. S.; FERNANDES, E. Mapeamento de processos como instrumento de apoio à implementação do custeio baseado em atividades nas organizações. *Encontro Nacional de Engenharia de Produção, Florianópolis*, p. 2099–2106, nov. 2004. Disponível em: <http://www.abepro.org.br/biblioteca/ENEGEP2004_Enegep0302_0606.pdf>. Acesso em: 01 nov. 2016.

[10] THIOLLENT, M. *Metodologia da pesquisa-ação.* [S.l.]: Cortez, 2011.

[11] MILAN, G. S.; VERSETTI, R. Melhorias em processos com impacto na eficiência operacional: um estudo ambientado em um laboratório de análises clínicas. *Revista Produção Online*, v. 12, n. 4, p. 1031–1056, 2012.

[12] DENZIN, N. K. *The research act: A theoretical introduction to sociological methods*. [S.l.]: Transaction publishers, 1973.

[13] MINAYO, M. C. de S. *Pesquisa social: teoria, método e criatividade*. [S.l.]: Editora Vozes Limitada, 2011.

[14] HUERTA, M. et al. Nih working definition of bioinformatics and computational biology. *US National Institute of Health*, 2000.

[15] ZOMAYA, A. Y. *Parallel computing for bioinformatics and computational biology: models, enabling technologies, and case studies*. [S.l.]: John Wiley & Sons, 2006.

[16] VERLI, H. *Bioinformática: da Biologia á Flexibilidade Molecular*. [S.l.]: SBBq, 2014.

[17] BORGES-OSÓRIO, M. R.; ROBINSON, W. M. *Genética Humana 3ed*. [S.l.]: Artmed Editora, 2013.

[18] LEE, W. et al. *From X-rays to DNA: How Engineering Drives Biology*. [S.l.]: MIT Press, 2013.

[19] HAGEN, J. B. The origins of bioinformatics. *Nature Reviews Genetics*, Nature Publishing Group, v. 1, n. 3, p. 231–236, 2000.

[20] HAUBEN, R. *The Birth and Development of the ARPANET*. [S.l.]: Netizens, 1995.

[21] MOUNT, D. W. Bioinformatics: Sequence and genome analysis. *Cold Spring Harbour Laboratory Press*, 2004.

[22] SILVA, J. M. e. *Bioquímica da Informação Genética*. [S.l.]: Publicações Ciência e Vida, Ltda, 2006.

[23] SANGER, F. et al. Nucleotide sequence of bacteriophage phi x174 dna. *Nature*, v. 265, p. 687–695, fev. 1977.

[24] COOK-DEEGAN, R. M. The human genome project: the formation of federal policies in the united states, 1986–1990. *Biomedical politics*, Washington, DC: National Academy Press, p. 99–167, 1991.

[25] KHAN, F. A. *Biotechnology in Medical Sciences*. [S.l.]: CRC Press, 2014.

[26] BAIROCH, A.; APWEILER, R. The swiss-prot protein sequence data bank and its new supplement trembl. *Nucleic Acids Research*, Oxford Univ Press, v. 24, n. 1, p. 21–25, 1996.

[27] COLLEN, M. F. *Computer medical databases: the first six decades (1950–2010)*. [S.l.]: Springer Science & Business Media, 2011.

[28] CASTRO, F. de. Uma década de bioinformática. *Agência FAPESP*, 2009. Disponível em: <agencia.fapesp.br/uma_decada_de_bioinformatica/10344>. Acesso em: 27 jul. 2016.

[29] DILLENBURG, D. Relatório do grupo de trabalho da área de ciências exatas e da terra. *INFOCAPES: boletim informativo*, CAPES, v. 9, n. 2-3, p. 17–23, abr./set. 2001.

[30] CASTRO, C. d. M. A pós-graduação e a capes: o que deu certo e o que deve ser feito. *INFOCAPES: Boletim Informativo*, CAPES, v. 10, n. 2, p. 138–149, abr./jun. 2002.

[31] REDE Ampliada. *Pesquisa Fapesp*, v. 71, p. 21, jan. 2002. Disponível em: <http://revistapesquisa.fapesp.br/2002/01/01/rede-ampliada/>. Acesso em: 6 ago. 2016.

[32] LNCC. Inauguração da unidade de genômica computacional darcy fontoura de almeida. 2008. Disponível em: <http://www.lncc.br/eventoSeminario/eventoconsultar.php?idt_evento=599>. Acesso em: 27 jul. 2016.

[33] LNCC. Lncc instala em sua sede o maior computador da américa latina. 2015. Disponível em: <http://www.lncc.br/lncc/supercomputador.php?idt_noticia=988>. Acesso em: 27 jul. 2016.

[34] SAKHARKAR, M.; KANGUEANE, P.; MATHURA, V. S. Biological sequence databases. In: *Bioinformatics: A Concept-Based Introduction*. [S.l.]: Springer, 2009. p. 39–46.

[35] ARAÚJO, N. D. d. et al. A era da bioinformática: seu potencial e suas implicações para as ciências da saúde. *Estudos de biologia*, v. 30, n. 70-72, p. 143–148, 2008.

[36] METZKER, M. L. Sequencing technologies—the next generation. *Nature reviews genetics*, Nature Publishing Group, v. 11, n. 1, p. 31–46, 2010.

[37] HODKINSON, B. P.; GRICE, E. A. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in wound care*, Mary Ann Liebert, Inc., v. 4, n. 1, p. 50–58, 2015.

[38] MELLET, J.; GRAY, C. M.; PEPPER, M. S. Hla typing: Conventional techniques v. next-generation sequencing. *SAMJ: South African Medical Journal*, v. 106, n. 1, p. 88–91, 2016.

[39] MORI, A. et al. Next generation sequencing: new tools in immunology and hematology. *Blood research*, v. 48, n. 4, p. 242–249, 2013.

[40] CHOO, S. Y. The hla system: genetics, immunology, clinical testing, and clinical implications. *Yonsei medical journal*, v. 48, n. 1, p. 11–23, 2007.

[41] REDOME. *O Que é o Transplante de Medula Óssea*. Disponível em: <http://redome.inca.gov.br/medula-ossea/o-que-e-o-transplante-de-medula-ossea/>. Acesso em: 10 jan. 2017.

[42] REDOME. *O que é o REDOME*. Disponível em: <http://redome.inca.gov.br/faqs/o-que-e-o-redome/>. Acesso em: 06 mar. 2017.

[43] REDOME. *O Que Faz*. Disponível em: <http://redome.inca.gov.br/o-redome/o-que-faz/>. Acesso em: 06 mar. 2017.

[44] REDOME. *Dados*. Disponível em: <http://redome.inca.gov.br/o-redome/dados/>. Acesso em: 06 mar. 2017.

[45] REDOME. *Transplante de Medula Óssea*. Disponível em: <http://redome.inca.gov.br/paciente/transplante-de-medula-ossea/>. Acesso em: 06 mar. 2017.

[46] BURKE, M. D. Laboratory medicine in the 21st century. *American journal of clinical pathology*, The Oxford University Press, v. 114, n. 6, p. 841–846, 2000.

[47] IFCC-WASPALM. *Principles of Clinical Laboratory Accreditation*. 1999. Disponível em: <http://www.waspalm.org/content/education/laboratory-accreditation>. Acesso em: 29 out. 2016.

[48] CAMPANA, G. A.; OPLUSTIL, C. P.; FARO, L. B. d. Tendências em medicina laboratorial. *J. bras. patol. med. lab*, v. 47, n. 4, p. 399–408, 2011.

[49] MOUMTZOGLOU, A. *Laboratory Management Information Systems: Current Requirements and Future Perspectives*. [S.l.]: IGI Global, 2014.

[50] LIMA-OLIVEIRA, G. S. et al. Controle da qualidade na coleta do espécime diagnóstico sanguíneo: iluminando uma fase escura de erros pré-analíticos. *Jornal Brasileiro de Patologia e Medicina Laboratorial*, v. 45, n. 6, p. 441–447, 2009.

[51] CAMPANA, G. A.; OPLUSTIL, C. P. Conceitos de automação na medicina laboratorial: revisão de literatura. *J. bras. patol. med. lab*, v. 47, n. 2, p. 119–127, 2011.

[52] FRIEDMAN, B. A. The total laboratory solution: a new laboratory e-business model based on a vertical laboratory meta-network. *Clinical chemistry*, v. 47, n. 8, p. 1526–1535, 2001.

[53] FGV-EAESP-GVCIA. *27ª Pesquisa Anual do uso de TI*. 2016. Disponível em: <http://eaesp.fgvsp.br/ensinoeconhecimento/centros/cia/pesquisa>. Acesso em: 05 nov. 2016.

[54] UNILAB. *A importância da tecnologia para otimização dos processos no laboratório*. 2015. Disponível em: <http://www.unilab.com.br/materiais-educativos/artigos/tecnologia/a-importancia-da-tecnologia-para-otimizacao-dos-processos-no-laboratorio/>. Acesso em: 05 nov. 2016.

[55] HUNT, V. D. *Process mapping: how to reengineer your business processes*. [S.l.]: John Wiley & Sons, 1996.

[56] PEREIRA, N. F. et al. Seleção de doador de medula óssea ou sangue periférico. *Revista Brasileira de Hematologia e Hemoterapia*, SciELO Brasil, v. 32, p. 03–05, 2010. Disponível em: <http://www.scielo.br/pdf/rbhh/v32s1/aop22010.pdf>. Acesso em: 06 mar. 2017.

[57] KAHL, G. *The dictionary of genomics, transcriptomics and proteomics*. [S.l.]: John Wiley & Sons, 2015.

[58] ILLUMINA. *HiSeq® 2500 Sequencing System*. 2015. Disponível em: <http://www.illumina.com/content/dam/illumina-marketing/documents/products/datasheets/datasheet_hiseq2500.pdf>. Acesso em: 09 dez. 2016.

[59] ILLUMINA. *HiSeq® 2500 System Guide*. 2015. Disponível em: <http://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/hiseq2500/hiseq2500-systemguide15035786-01.pdf>. Acesso em: 09 dez. 2016.

[60] COCK, P. J. et al. The sanger fastq file format for sequences with quality scores, and the solexa/illumina fastq variants. *Nucleic acids research*, Oxford Univ Press, v. 38, n. 6, p. 1767–1771, 2010.

[61] ILLUMINA. *BaseSpace User Guide*. 2014. Disponível em: <http://support.illumina.com/content/dam/illumina-support/documents/documentation/software_documentation/basespace/basespace-userguide15044182e.pdf>. Acesso em: 09 dez. 2016.

[62] VALLE, R.; OLIVEIRA, S. B. d. *Análise e modelagem de processos de negócios: Foco na notação BPMN (Business Process Modeling Notation)*. [S.l.]: Atlas, 2011.

[63] BALTZAN, P.; PHILLIPS, A. *Sistemas de Informação*. [S.l.]: McGraw Hill Brasil, 2012.

[64] E-WORKFLOW. *The Workflow Portal*. Disponível em: <http://www.e-workflow.org/>. Acesso em: 27 fev. 2017.

[65] HARTL, M. *Ruby on rails tutorial: learn Web development with rails*. Addison-Wesley Professional, 2016. Disponível em: <https://www.railstutorial.org/book/beginning#cha-beginning>. Acesso em: 23 dez. 2016.

[66] NERY, A. S. et al. *Guia de Referência Rápida Ruby on Rails - MVC - Parte 1*.

[67] CAELUM. *Desenvolvimento Ágil para Web com Ruby on Rails*. Disponível em: <https://www.caelum.com.br/apostila-ruby-on-rails/>. Acesso em: 26 fev. 2017.

[68] NERY, A. S. et al. *Guia de Referência Rápida Ruby on Rails - MVC - Parte 2*.

[69] TRIVEDE, K. S. et al. Modeling high availability. In: IEEE COMPUTER SOCIETY. *Proceedings of the 12th Pacific Rim International Symposium on Dependable Computing*. 2006. p. 154–164. Disponível em: <http://amod.ee.duke.edu/PAPERS/sunprdc06.pdf>. Acesso em: 06 mar. 2017.

[70] APACHE. *The Apache Software Foundation*. Disponível em: <https://www.apache.org/>. Acesso em: 07 mar. 2017.

[71] PASSENGER. *Deploying a Ruby app with Passenger to production*. Disponível em: <https://www.phusionpassenger.com/library/walkthroughs/deploy/ruby/>. Acesso em: 07 mar. 2017.

[72] POSTGRESQL. *PostgreSQL 9.2.20 Documentation*. Disponível em: <https://www.postgresql.org/docs/9.2/static/index.html>. Acesso em: 07 mar. 2017.

[73] DRBD. *High Availability with DRBD*. Disponível em: <https://www.linbit.com/en/products-and-services/drbd/>. Acesso em: 07 mar. 2017.

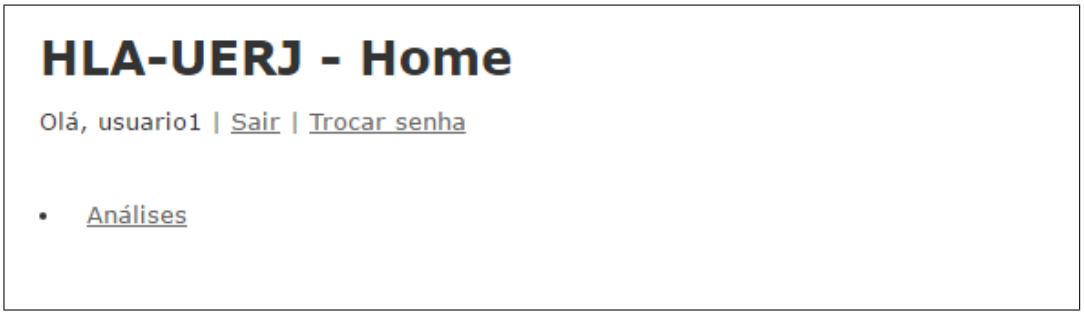
APÊNDICE - Telas do Protótipo

Com o objetivo de exemplificar e iniciar o projeto do sistema proposto, foi desenvolvido um protótipo com as funcionalidades propostas de autenticação de multiusuários, de cadastro de tipos e fases e de cadastro de análises. Para o desenvolvimento do protótipo foi utilizado o sistema operacional Linux distribuição Ubuntu 14.04, a versão 5.0.1 do Rails, a versão 2.0 do Ruby, o banco de dados SQLite3, que é o banco padrão do Rails e as gems paperclip-5.0.0, cancancan-1.16.0, devise-4.2.0 e load_and_authorize_resource-0.4.0. Seguem as telas desenvolvidas:

Figura 31 - Tela de Login do protótipo

Fonte: O autor, 2017.

Figura 32 - Tela Inicial do protótipo para o usuário comum

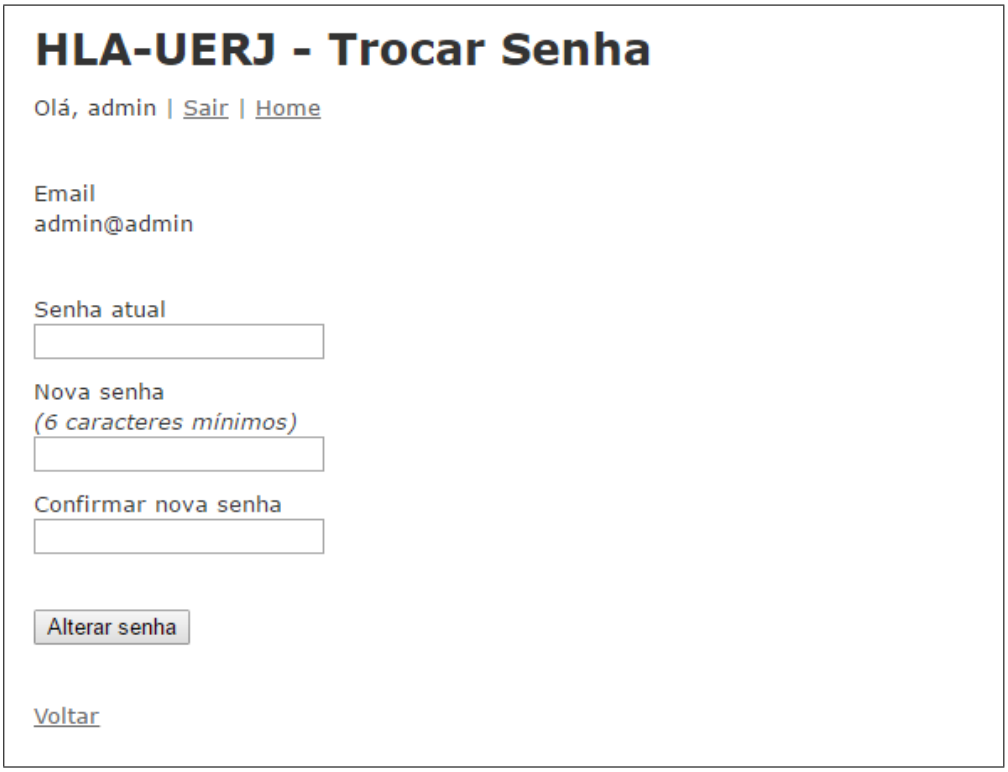


Fonte: O autor, 2017.

Figura 33 - Tela Inicial do protótipo para o usuário administrador

Fonte: O autor, 2017.

Figura 34 - Tela do protótipo para trocar a própria senha



Fonte: O autor, 2017.

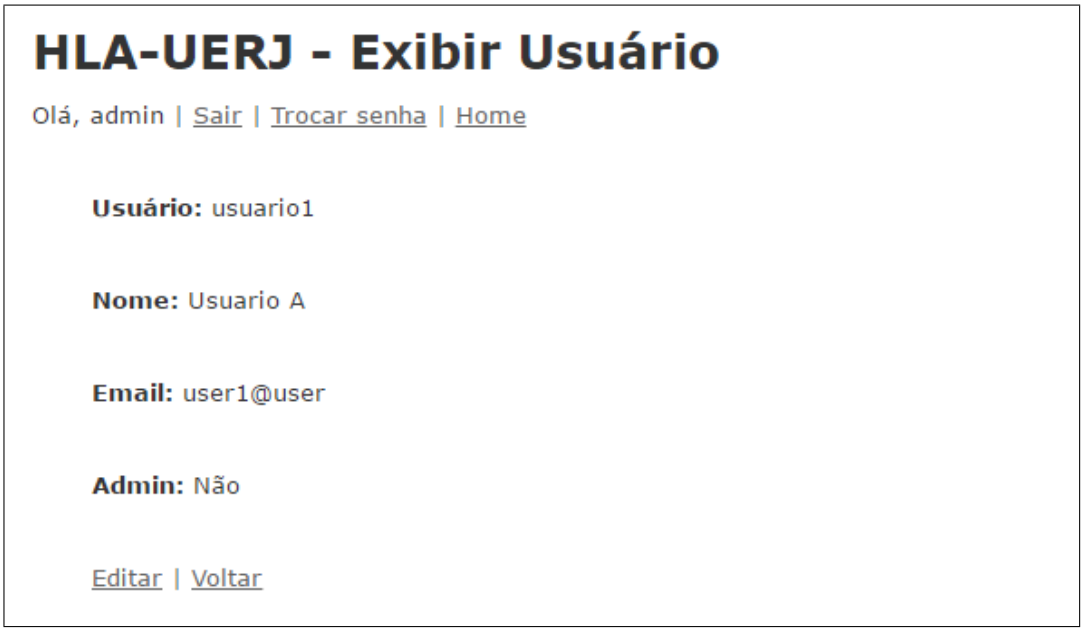
Figura 35 - Tela do protótipo para listar usuários

Fonte: O autor, 2017.

Figura 36 - Tela do protótipo para criar novo usuário

Fonte: O autor, 2017.

Figura 37 - Tela do protótipo para exibir usuário



Fonte: O autor, 2017.

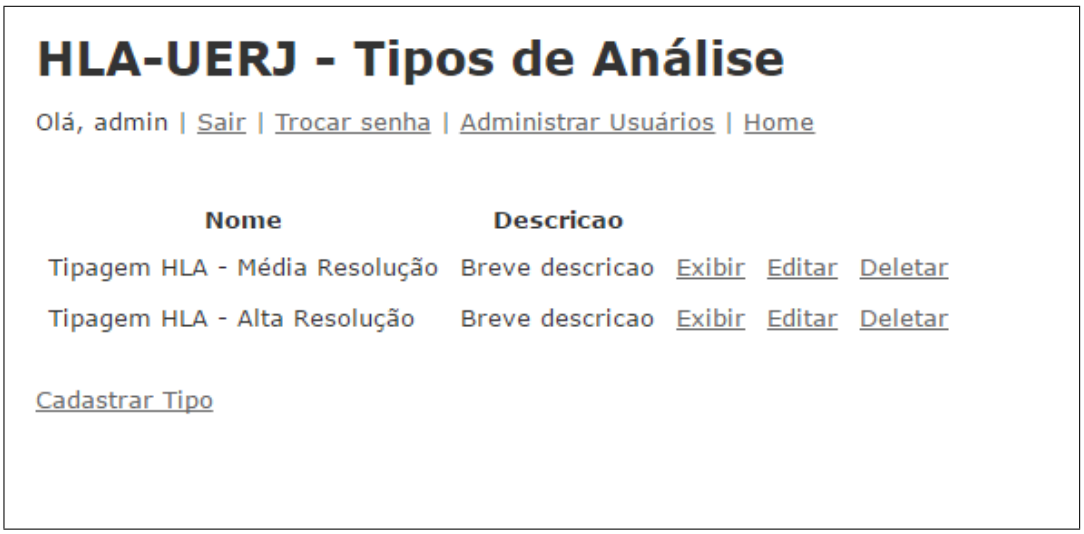
Figura 38 - Tela do protótipo para editar usuário

Fonte: O autor, 2017.

Figura 39 - Tela do protótipo para trocar a senha do usuário

Fonte: O autor, 2017.

Figura 40 - Tela do protótipo para listar tipos de análise

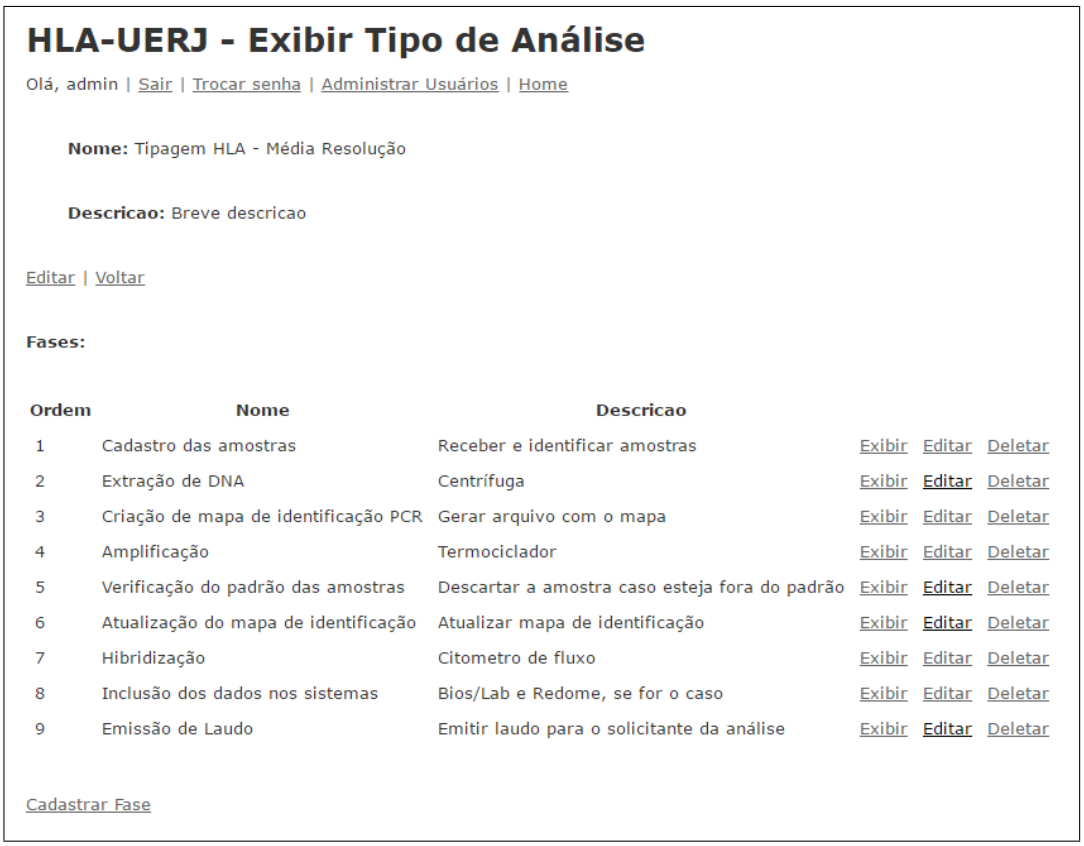


Fonte: O autor, 2017.

Figura 41 - Tela do protótipo para criar tipos de análise

Fonte: O autor, 2017.

Figura 42 - Tela do protótipo para exibir tipos de análise

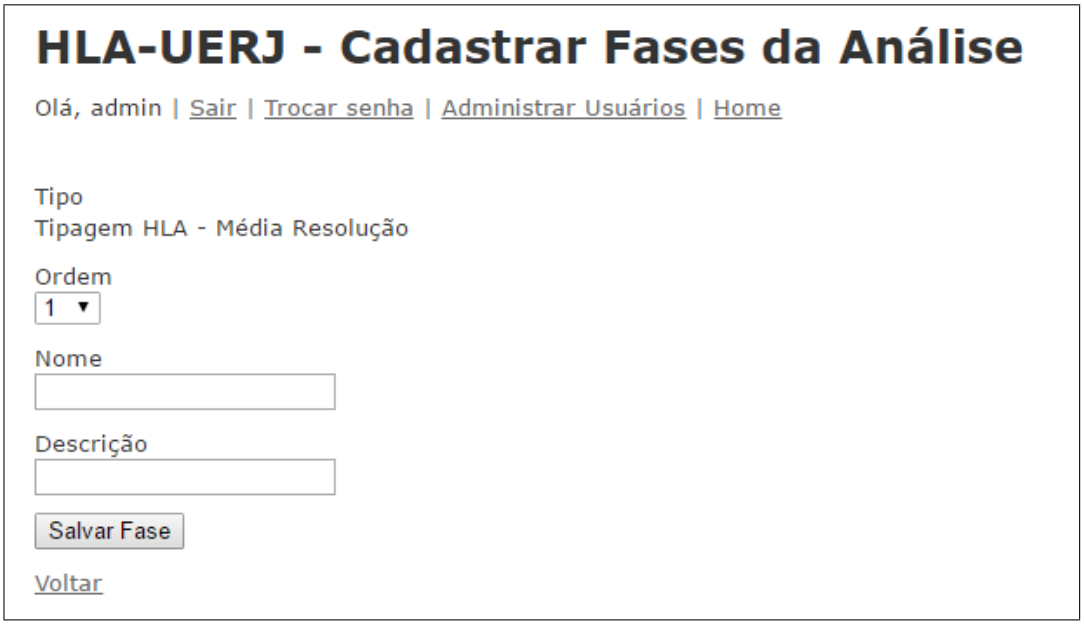


Fonte: O autor, 2017.

Figura 43 - Tela do protótipo para editar tipos de análise

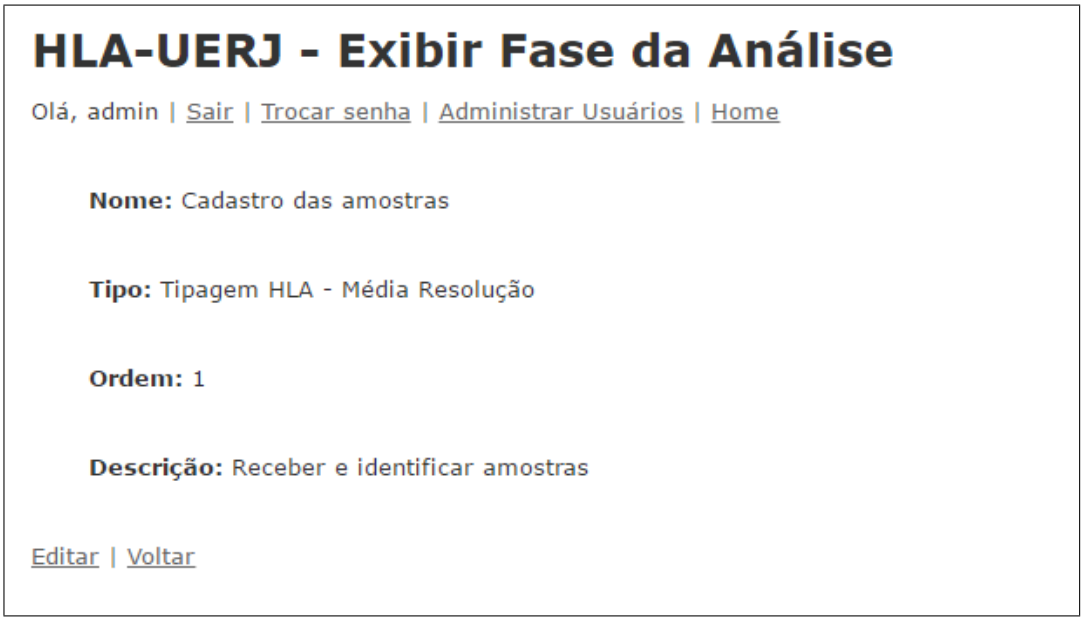
Fonte: O autor, 2017.

Figura 44 - Tela do protótipo para criar fase



Fonte: O autor, 2017.

Figura 45 - Tela do protótipo para exibir fase



Fonte: O autor, 2017.

Figura 46 - Tela do protótipo para editar fase

Fonte: O autor, 2017.

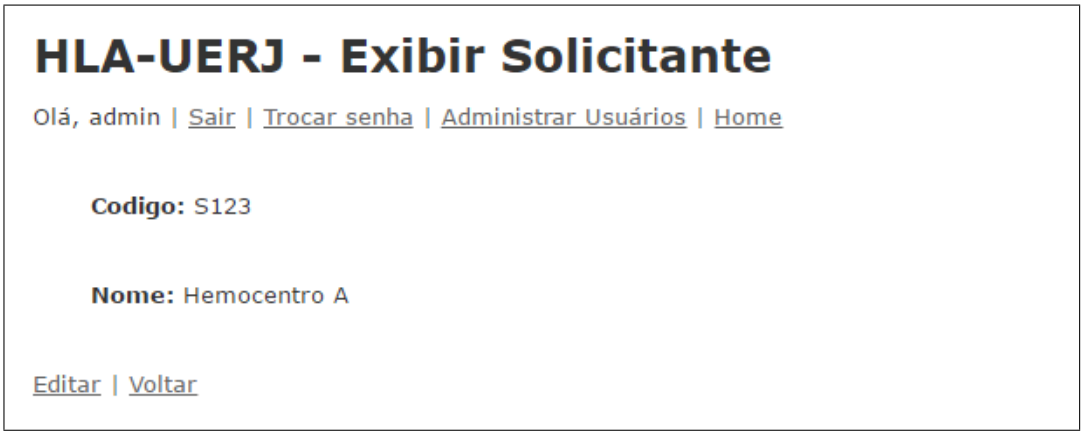
Figura 47 - Tela do protótipo para listar solicitantes

Fonte: O autor, 2017.

Figura 48 - Tela do protótipo para criar solicitantes

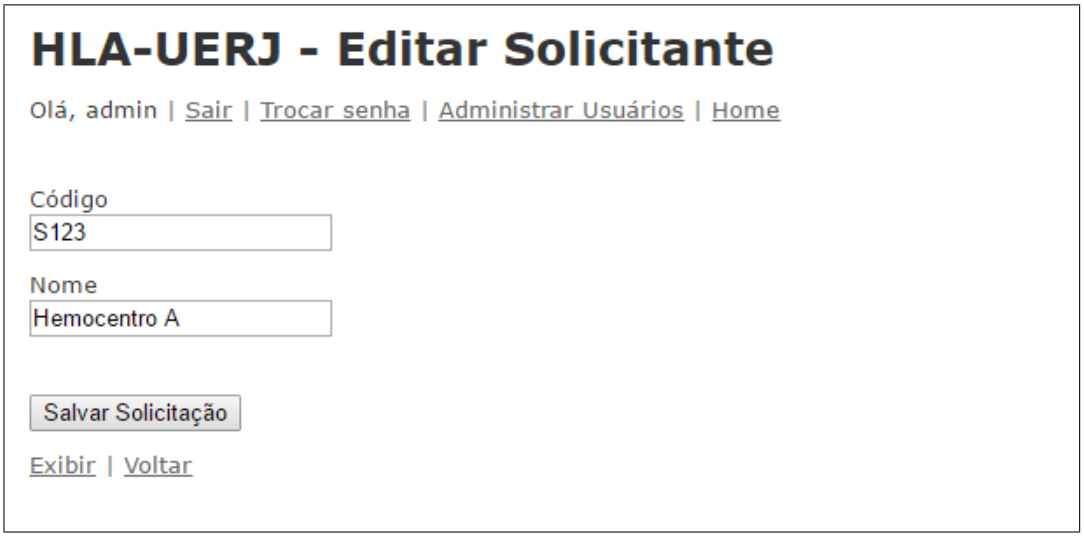
Fonte: O autor, 2017.

Figura 49 - Tela do protótipo para exibir solicitantes



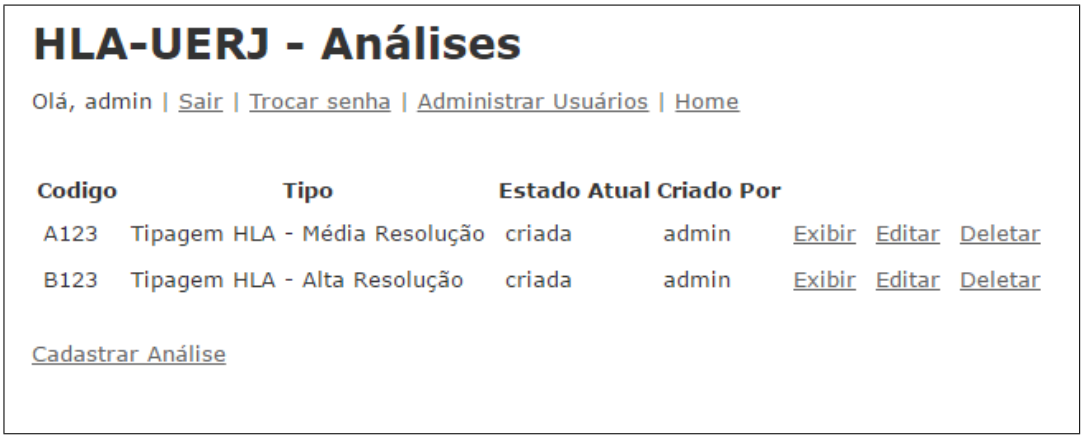
Fonte: O autor, 2017.

Figura 50 - Tela do protótipo para editar solicitantes



Fonte: O autor, 2017.

Figura 51 - Tela do protótipo para listar análises

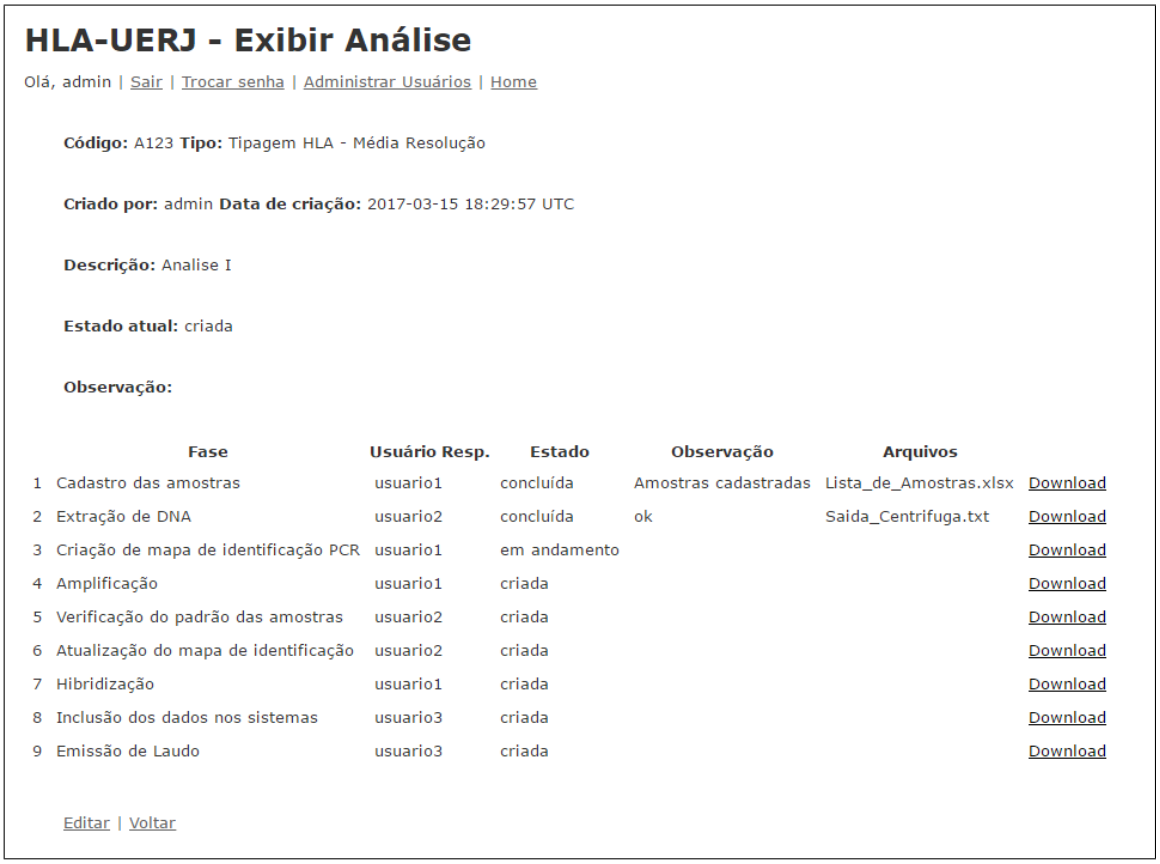


Fonte: O autor, 2017.

Figura 52 - Tela do protótipo para criar análise

Fonte: O autor, 2017.

Figura 53 - Tela do protótipo para exibir análise



HLA-UERJ - Exibir Análise

Olá, admin | Sair | Trocar senha | Administrar Usuários | Home

Código: A123 **Tipo:** Tipagem HLA - Média Resolução

Criado por: admin **Data de criação:** 2017-03-15 18:29:57 UTC

Descrição: Analise I

Estado atual: criada

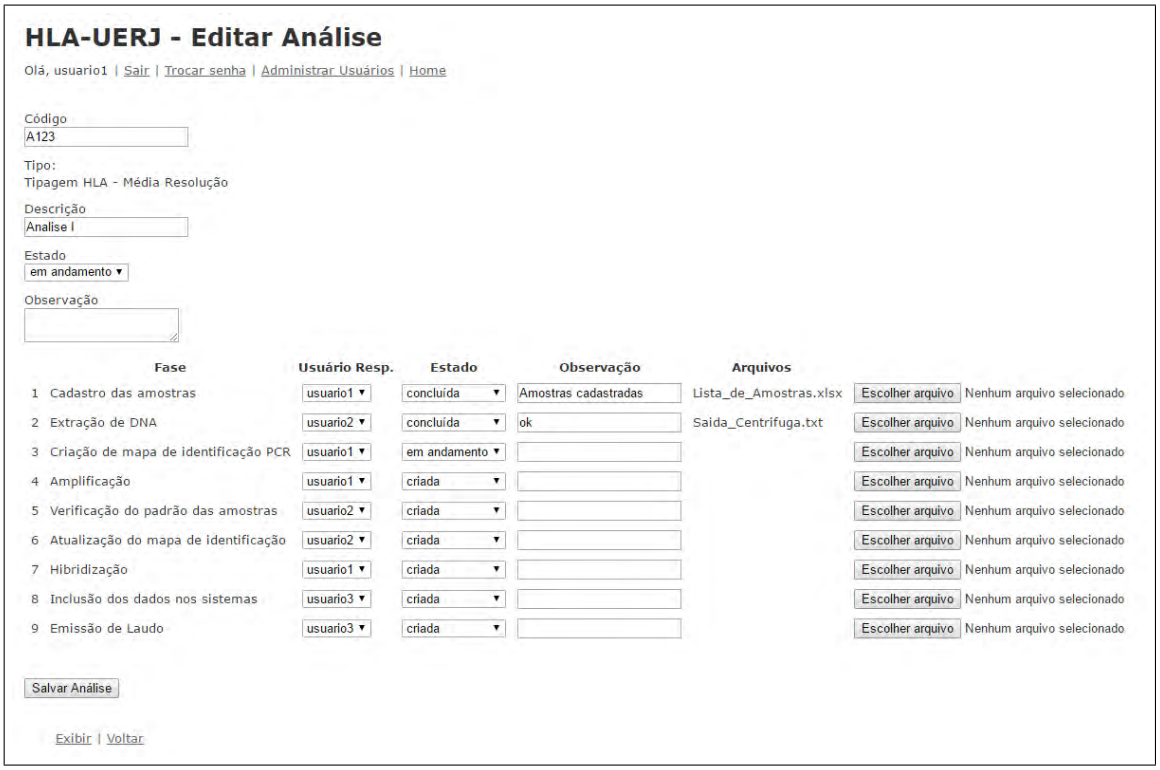
Observação:

	Fase	Usuário Resp.	Estado	Observação	Arquivos	
1	Cadastro das amostras	usuario1	concluída	Amostras cadastradas	Lista_de_Amostras.xlsx	Download
2	Extração de DNA	usuario2	concluída	ok	Saida_Centrifuga.txt	Download
3	Criação de mapa de identificação PCR	usuario1	em andamento			Download
4	Amplificação	usuario1	criada			Download
5	Verificação do padrão das amostras	usuario2	criada			Download
6	Atualização do mapa de identificação	usuario2	criada			Download
7	Hibridização	usuario1	criada			Download
8	Inclusão dos dados nos sistemas	usuario3	criada			Download
9	Emissão de Laudo	usuario3	criada			Download

Editar | Voltar

Fonte: O autor, 2017.

Figura 54 - Tela do protótipo para editar análise



Fonte: O autor, 2017.