

UNIVERSIDADE FEDERAL DO PARANÁ

ALLAN COSTA VIEIRA

FERNANDO DE OLIVEIRA MELO

BASE DE DADOS - FUNGOS PATOLÓGICOS

CURITIBA
2016

ALLAN COSTA VIEIRA
FERNANDO DE OLIVIEIRA MELO

BASE DE DADOS - FUNGOS PATOLÓGICOS

Trabalho de conclusão de curso apresentado
como requisito à conclusão do curso de
Tecnologia em Análise e Desenvolvimento de
Sistemas, Setor de Educação Profissional e
Tecnologia, Universidade Federal do Paraná.
Orientador: Jeroniza Nunes Marchaukoski

CURITIBA
2016

RESUMO

Arquivos GBK, GFF e Fasta guardam as mais diversas informações genéticas de um organismo, sendo amplamente utilizados na Bioinformática, mas apesar de poderem possuir a mesma informação, cada formato de arquivo as armazena de forma diferente. Este trabalho apresenta um sistema web capaz de guardar e converter os arquivos do formato GBK para os formatos GFF e Fasta, assim como disponibilizá-los para download. O sistema também funcionará como uma plataforma de trabalho, facilitando o armazenamento e visualização de arquivos genéticos e possibilitando que o andamento dos projetos seja facilmente consultado, arquivos disponibilizados e usuários gerenciados de acordo com uma hierarquia estabelecida.

Palavra Chave: Bioinformática, Arquivos genéticos, Visualização de dados biológicos, GFF, GBK

ABSTRACT

GBK, GFF and Fasta files store the most diverse genetic information of an organism, being widely used in bioinformatics, but despite the fact that they can have the same information, each file format has different ways of storing. This paper presents a web system capable of store and convert files from GBK to GFF and Fasta, and offers them for download. The system will also function as a work platform, facilitating the storage and visualization of genetic files and enabling the progress of each project to be easily consulted, files made available and users managed in accordance with an established hierarchy.

Key-word: Bioinformatics, Genetic files, Visualization of biological data, GFF, GBK

LISTA DE FIGURAS

FIGURA 1 – TELA INICIAL PÚBLICO.....	21
FIGURA 2 – TELA PROJETOS.....	22
FIGURA 3 – TELA ARQUIVOS.....	22
FIGURA 4 – TELA SOLICITAR ACESSO DE CONVIDADO.....	23
FIGURA 5 – TELA CADASTRO PESQUISADOR.....	23
FIGURA 6 – TELA SOLICITAR ACESSO DE COORDENADOR.....	24
FIGURA 7 – TELA INICIAL ADMINISTRADOR.....	25
FIGURA 8 – TELA EDITAR PROJETOS.....	25
FIGURA 9 – TELA ACEITAR PROJETOS E NOVOS COORDENADORES.....	26
FIGURA 10 – TELA EDITAR/ EXCLUIR COORDENADORES.....	26
FIGURA 11 – TELA INICIAL COORDENADOR.....	27
FIGURA 12 – TELA EDITAR/EXCLUIR PROJETOS.....	27
FIGURA 13 – TELA EDITAR/EXCLUIR ORGANISMOS.....	28
FIGURA 14 – TELA INCLUIR ARQUIVOS DE PROJETO.....	28
FIGURA 15 – TELA SUBMETER UM NOVO PROJETO.....	29
FIGURA 16 – TELA CADASTRA EQUIPE.....	29
FIGURA 17 – TELA ACEITAR PESQUISADORES/CONVIDADOS.....	30
FIGURA 18 – TELA EXCLUIR PESQUISADORES.....	30
FIGURA 19 – TELA EDITAR EQUIPES.....	31
FIGURA 20 – DIAGRAMA DE CASOS DE USO.....	
FIGURA 21 – DIAGRAMA ENTIDADE RELACIONAMENTO.....	
FIGURA 22 – DIAGRAMA DE TELAS – ACESSO PÚBLICO.....	
FIGURA 23 – DIAGRAMA DE TELAS – ACESSO DE ADMINISTRADOR.....	
FIGURA 24 – DIAGRAMA DE TELAS – ACESSO DE COORDENADOR.....	
FIGURA 25 – DIAGRAMA DE TELAS – ACESSO DE PESQUISADOR.....	
FIGURA 26 – DIAGRAMA DE CLASSES.....	

LISTA DE SIGLAS

DNA	-Ácido desoxirribonucleico
FASTA	-Formato utilizado para armazenar sequências de bases e de aminoácidos em arquivo texto
GenBank	-Banco de dados público do National Center for Biology Information
GBK	-GenBank File
GFF	-General Feature Format
HTML	-HyperText Markup Language
IDE	-Integrated Development Environment
MVC	-Model-View-Controller
NCBI	-National Center for Biotechnology Information
NR	-Banco de dados não redundante disponibilizado pelo NCBI
PPG-BIOINFO	- Programa de Pós-graduação em Bioinformática
RNA	-Ácido ribonucleico
RUP	-Rational Unified Process
SGBD	-Sistema Gerenciador de Bancos de Dados
UML	-Unified Modeling Language
WBS	-Work Breakdown Structure
XML	-Extensible Markup Language

LISTA DE TABELAS

TABELA 1 – PLANO DE RISCOS.....	16
TABELA 2 – DICIONÁRIO DE DADOS.....	

SUMÁRIO

1 INTRODUÇÃO	9
1.1 OBJETIVOS	10
1.2 OBJETIVOS ESPECÍFICOS.....	10
2 FERRAMENTAS UTILIZADAS	12
2.6 POSTGRESQL.....	13
3 METODOLOGIA.....	14
3.1 MODELO DE PROCESSO DE ENGENHARIA DE SOFTWARE.....	14
3.2 PLANO DE RISCOS.....	15
3.3 MATERIAIS	16
3.3.1 Ambiente de Desenvolvimento	16
3.3.2 Linguagem de Programação	16
3.3.4 Banco de Dados.....	17
3.4.5 Ferramenta para Modelagem UML	17
3.4.6 Ferramenta para Criação do Diagrama Entidade Relacionamento.....	17
3.5 FERRAMENTA PARA CONTROLE DE VERSÃO	17
3.6 RUBYGEMS	18
4 DESENVOLVIMENTO DO PROJETO	19
5 APRESENTAÇÃO DO SOFTWARE	21
5.5 INSTALAÇÃO	36
6 CONSIDERAÇÕES FINAIS.....	39
REFERÊNCIAS	42
APÊNDICE A - DIAGRAMA DE CASOS DE USO	45
APÊNDICE B - DIAGRAMA ENTIDADE RELACIONAMENTO	46
APÊNDICE C- DICIONÁRIO DE DADOS.....	47
APÊNDICE D- DIAGRAMA DE TELAS.....	50
APÊNDICE D- DIAGRAMA DE CLASSES.....	54

1 INTRODUÇÃO

O Programa de Pós-graduação em Bioinformática de Universidade Federal do Paraná foi criado em 2009 com intenção de atender a grande demanda por profissionais capazes de convergir entre os estudos de biologia molecular e computação, aprofundando-se cada vez mais no campo interdisciplinar chamado Bioinformática.

A Bioinformática é uma ciência que atua na manipulação e extração de conhecimento de dados biológicos. O crescimento da área deve-se à necessidade de manipular quantidades enormes de dados genéticos e bioquímicos, utilizando recursos computacionais para catalogá-las, organizá-las e estruturá-las (RASHIDI, 2000).

O PPG-BIOINFO necessita cada vez mais de ferramentas que facilitem o trabalho de pesquisa e armazenamento das informações trabalhadas, e, visando criar um banco de dados para armazenamento das informações genéticas de fungos patológicos, esperamos contribuir de maneira efetiva para que o trabalho de todo o grupo de pesquisa seja facilitado.

Nossa plataforma pretende trazer para o grupo, além de um local para o novo local para armazenamento de informações, uma nova forma de organização de trabalho, possibilitando assim um maior aproveitamento do tempo de seus colaboradores.

1.1 OBJETIVOS

Desenvolver uma plataforma web capaz de armazenar arquivos genéticos, organizando-os de acordo com o organismo estudado e o projeto em que se encontra. Além disso, fazer com que essa plataforma tenha um sistema de gerenciamento hierarquico de usuários, possibilitando seguir os moldes organizacionais do PPG-BIOINFO.

1.2 OBJETIVOS ESPECÍFICOS

Este trabalho tem os seguintes objetivos específicos:

- Possibilitar ao sistema fazer upload de arquivos GBK
- Possibilitar ao sistema fazer a conversão de arquivos GBK para GFF e Fasta
- Possibilitar ao sistema armazenar esses arquivos de forma compactada, afim de reduzir o espaço ocupado no banco de dados e possibilitar que os downloads sejam feitos de forma mais rápida pelos usuários
- Possibilitar ao visitante o acesso a dados públicos
- Possibilitar ao visitante o download de arquivos genéticos que tenham sido definidos como públicos
- Possibilitar ao visitante solicitar acesso de coordenador para aprovação de um administrador
- Possibilitar ao visistante solicitar acesso de pesquisador a um projeto para aprovação de um coordenador
- Possibilitar ao administrador do sistema aceitar ou rejeitar solicitações públicas de acesso como coordenador
- Possibilitar ao administrador redesignar um projeto a outro coordenador
- Possibilitar ao administrador editar e excluir coordenadores
- Possibilitar ao coordenador cadastrar, editar e excluir um projeto
- Possibilitar ao coordenador editar uma equipe
- Possibilitar ao coordenador aceitar ou rejeitar solicitação de pesquisadores ou convidados para um projeto
- Possibilitar ao coordenador editar um pesquisador
- Possibilitar ao coordenador incluir, editar e excluir um organismo
- Possibilitar ao coordenador fazer upload de arquivos genéticos de um organismo

- Possibilitar ao coordenador o download de arquivos genéticos de seus projetos
- Possibilitar ao pesquisador o acesso aos dados do projeto em que faz parte
- Possibilitar ao pesquisador o download de arquivos genéticos de seus projetos
- Possibilitar ao pesquisador solicitar acesso de convidado a um projeto de um outro coordenador

2 FERRAMENTAS UTILIZADAS

O objetivo deste capítulo é apresentar os recursos que foram utilizados para o desenvolvimento do projeto.

2.1 BALSAMIQ MOCKUPS

Para a criação da estrutura do wireframe das telas, foi utilizado o software Balsamiq Mockups. Com ele, conseguiu-se uma forma ágil e realista para o desenvolvimento do layout modelo, tudo isso devido às facilidades trazidas pela ferramenta (BALSAMIQ, 2016).

2.2 UML

A UML (*Unified Modeling Language*) é uma linguagem visual utilizada para modelar software baseados no paradigma de orientação a objetos. É uma linguagem de modelagem de propósito geral que pode ser aplicada a todos os domínios de aplicação. Essa linguagem tornou-se, nos últimos anos, a linguagem-padrão de modelagem adotada internacionalmente pela indústria de engenharia de software (GUEDES, 2011).

Neste trabalho utilizamos o diagrama de caso de uso, o diagramas de classe e o diagrama de telas. O diagrama de caso de uso permite organizar os requisitos do sistema dando uma visão geral. O diagrama de classes representa a estrutura do sistema orientado a objetos e o diagrama de telas apresenta o fluxo das atividades do aplicativo.

2.3 ASTAH PROFESSIONAL

Para a criação dos diagramas UML utilizamos o Astah, que é uma ferramenta para específica para a criação desses diagramas. O Astah está a mais de 10 anos disponível no mercado (primeiramente com o nome JUDE) e com ele é possível criar todos os diagramas de uma forma simples, intuitiva e padronizada.

2.4 ORACLE VM VIRTUALBOX

A Virtual Machine da Oracle, VirtualBox, é um software de virtualização. Com ela é possível rodar os mais diversos sistemas operacionais sem a necessidade de modificação, possibilitando assim o desenvolvimento em um sistema Linux, permitindo termos um sistema mais parecido com o do PPG-BIOINFO do que o disponível em nossas máquinas (ORACLE, 2016).

2.5 MYSQL WORKBENCH

O MySQL Workbench é uma ferramenta para design de banco de dados, também de propriedade da Oracle. Além do design do banco de dados, única funcionalidade a qual utilizamos, a ferramenta integra uma plataforma para desenvolvimento em SQL, administração de bancos, monitoramento de performance e migração entre bancos (MYSQL, 2016).

2.6 POSTGRESQL

PostgreSQL é um SGBD que implementa um banco de dados SQL relacional. O PostgreSQL é um software livre de domínio público e multiplataforma. Este SGBD foi escolhido pela equipe pela necessidade do grupo de Pós-graduação em Bioinformática da UFPR em unificar o desenvolvimento de novas ferramentas sobre um único SGBD, facilitando assim a manutenção e o aprendizado (POSTGRESQL, 2016).

3 METODOLOGIA

A função deste capítulo é apresentar as metodologias e ferramentas que foram utilizadas para o desenvolvimento deste aplicativo. Também serão mostradas as responsabilidades dadas a cada membro.

3.1 MODELO DE PROCESSO DE ENGENHARIA DE SOFTWARE

Neste projeto foi utilizado o modelo RUP (*Rational Unified Process*) que fornece diretrizes para definir as tarefas e atribuir responsabilidades aos membros do projeto. O RUP possui quatro fases: iniciação, elaboração, construção e transição. Na iniciação define-se o escopo do projeto. Na elaboração é obtida uma visão abrangente do sistema, através da construção de protótipos e também é definida a arquitetura do sistema. Na construção, o foco está no desenvolvimento do sistema. Por fim, na transição, o produto é transferido ao usuário e o projeto é avaliado e pode ser concluído. Em todas as fases há o gerenciamento dos requisitos e dos recursos do projeto. Por isso o RUP é baseado no desenvolvimento iterativo, que por sua vez é mais flexível quanto às mudanças de escopo durante o desenvolvimento do projeto (IBM, 2014).

3.2 PLANO DE RISCOS

No plano de riscos foram definidos os impactos e a probabilidade de determinados eventos acontecerem, bem como as possíveis ações que poderiam ser tomadas. O plano de riscos do projeto pode ser observado na TABELA 1.

TABELA 1 – PLANO DE RISCOS

Nº	Risco	Consequência	Ação	Probabilidade	Impacto	Classificação
1	Falta de tempo para o desenvolvimento do aplicativo.	Impossibilidade de entregar ou atraso na entrega.	Divisão de tarefas e monitoramento do andamento.	Alto	Alto	7
2	Mudança de requisitos do aplicativo.	Necessidade de mudar o direcionamento do projeto.	Mudanças na estrutura do projeto.	Alto	Alto	7
3	Conhecimento limitado em bioinformática.	Dificuldade no desenvolvimento do projeto.	Estudo e pesquisa na internet, livros e em outros materiais da PPG-BIOINFO.	Alto	Moderado	6
4	Perda de dados.	Impossibilidade de continuar o projeto.	Manter o repositório do GIT atualizado	Moderado	Alto	6
5	Desistência de um membro da equipe.	Não haveria a possibilidade de entrega do trabalho, devido a resolução do curso.	Conseguir outro membro para o grupo.	Moderado	Alto	6
6	Infraestrutura insuficiente.	Falta de equipamentos para desenvolvimento e testes.	Utilizar emuladores para as necessidades específicas.	Moderado	Moderado	4

3.3 MATERIAIS

O objetivo deste capítulo é descrever as ferramentas utilizadas para o desenvolvimento do projeto.

3.3.1 Ambiente de Desenvolvimento

Para o desenvolvimento do aplicativo, foram utilizados dois computadores:

- Notebook CyberPower MS16-GC, processador Intel Core i7-4810MQ 2.80GHz, memória RAM 16GB, placa de vídeo Nvidia GTX 860M 4GB, sistema operacional Windows 10 Home;
 - Oracle VM Virtual Box, Ubuntu 16.04
- Notebook Dell Latitude 3450, processador Intel Core i5 5200U 2.20GHz, memória RAM 8GB, placa de vídeo Intel HD Graphics 5500, sistema operacional Windows 8.1 Pro;
 - Oracle VM Virtual Box, Ubuntu 16.04

3.3.2 Linguagem de Programação

A linguagem de programação escolhida pelos usuários foi Ruby. O Ruby é uma linguagem relativamente nova, criada em 1995 por Yuri Matsumoto, podendo ser base para diversos paradigmas de programação. Escolhemos trabalhar com orientação a objetos para poder aplicar boa parte do conhecimento adquirido em aula no curso de Análise e Desenvolvimento de Sistemas (RUBY, 2016).

A escolha dessa linguagem se deu pela necessidade do grupo de Pós-graduação em Bioinformática da UFPR em unificar o desenvolvimento de novas ferramentas sobre uma única linguagem, facilitando assim a manutenção e o aprendizado.

3.3.3 Framework de Desenvolvimento

Para um melhor uso da linguagem Ruby, utilizamos o framework Ruby on Rails durante o desenvolvimento do projeto.

Este framework facilita o desenvolvimento, implantação e manutenção, construindo uma aplicação web dentro da arquitetura MVC (RUBY ON RAILS, 2016).

3.3.4 Banco de Dados

Para implementação do banco de dados foi escolhido o PostgreSQL, pois é uma ferramenta de fácil uso, com bastante documentação disponível e bem conceituada no meio acadêmico e comercial, sendo já utilizado no PPG-BIOINFO em outros projetos.

3.4.5 Ferramenta para Modelagem UML

A ferramenta utilizada para criação dos diagramas UML foi o Astah Professional.

3.4.6 Ferramenta para Criação do Diagrama Entidade Relacionamento

A ferramenta utilizada para criação do diagrama entidade relacionamento foi MySQL Workbench, software gratuito para uso não comercial e desenvolvido pela Oracle. Essa ferramenta está disponível somente para o sistema operacional Windows.

3.5 FERRAMENTA PARA CONTROLE DE VERSÃO

O git é um sistema de controle de versão de softwares focado no desenvolvimento envolvendo um ou mais desenvolvedores. Ele também é um sistema de gerenciamento de código fonte, ou seja, todos os membros da equipe podem enviar alterações. Todos os arquivos e todo histórico são armazenados num repositório único (GIT, 2016).

Neste projeto foi usado o GitHub, que além de gerenciar as versões e as alterações no código, possui um repositório online, facilitando as alterações para os membros da equipe e o acesso a informações necessárias a qualquer momento.

3.6 RUBYGEMS

Gems são bibliotecas utilizadas para resolver diversos tipos de problema no Ruby. Os desenvolvedores utilizam as Gems para não precisarem criar uma solução já existente.

Através de pesquisas foi decidido pelos membros da equipe somente uma biblioteca externa seria utilizada no projeto, o BioRuby para poder converter os dados entre os mais diversos formatos necessários.

4 DESENVOLVIMENTO DO PROJETO

O projeto começou com a ideia de criação de um banco de dados capaz de armazenar o resultado de um parse de um arquivo genético com a intenção de ter as características de um organismo separadas e com isso não precisar recorrer ao arquivo completo todas as vezes que fosse necessário consultar uma informação. Porém, mudanças foram feitas no escopo ao decorrer do projeto e a ideia do parse foi deixada para uma implementação futura.

Com isso o projeto passou de um sistema de armazenamento de parsers para uma plataforma de trabalho, incluindo não somente as funcionalidades de download, conversão e upload de arquivos, como também o gerenciamento de usuários por hierarquia organizacional.

Dificuldades apareceram ao longo do projeto devido a pouca experiência dos desenvolvedores com o tipos de arquivos genéticos, essenciais para todo o projeto, o que acabou ocasionando dificuldade no entendimento das necessidades do PPG-BIOINFO assim como na definição do escopo final do projeto.

Boa parte dessa dificuldade veio da tentativa de conversão entre arquivos GBK, GFF e Fasta, mas pelo fato dos dois ultimos poderem não ter a mesma quantidade de informação que o arquivo GBK optou-se, depois de decorrido boa parte do projeto, somente pela conversão do arquivo GBK para os outros dois.

Um segundo problema relacionado aos arquivos genéticos era o seu tamanho. Podendo ocupar centenas de MB, armazená-los e transportá-los poderia ocupar bastante espaço e recurso computacional, por isso propusemos a compactação desses arquivos e implementamos essa funcionalidade. Devido o fato dos arquivos conterem diversas sequências de caracteres repetidos a compactação atinge números excelentes, possibilitando desperdiçar menos memória no banco de dados e recurso computacional para transporte.

Outro problema recorrente durante o desenvolvimento, principalmente na fase de implantação do projeto no servidor da PPG-BIOINFO, foi a queda, por vezes constante, da conexão com o servidor pela VPN. Diversas vezes foi necessário efetuar um retrabalho devido a essas quedas.

A divisão de tarefas foi mudando de acordo com as mudanças requisito e aproximação do prazo final. Apesar das diversas mudanças, a equipe conseguiu se adaptar a elas.

A equipe decidiu retirar os diagramas de sequência e substituí-los por diagramas de tela, pois além de não termos sequências tão grandes entre as classes os diagramas de telas mais fáceis de serem lidos e ilustram o que ocorre de forma mais simples e intuitiva.

5 APRESENTAÇÃO DO SOFTWARE

Neste capítulo será apresentada uma visão geral do software desenvolvido, mostrando suas telas e funcionalidades.

5.1 PERFIL PÚBLICO DE ACESSO

O acesso de todos usuários iniciasse pela tela da FIGURA 01, no qual é possível fazer login ou solicitar acesso de coordenador para o administrador

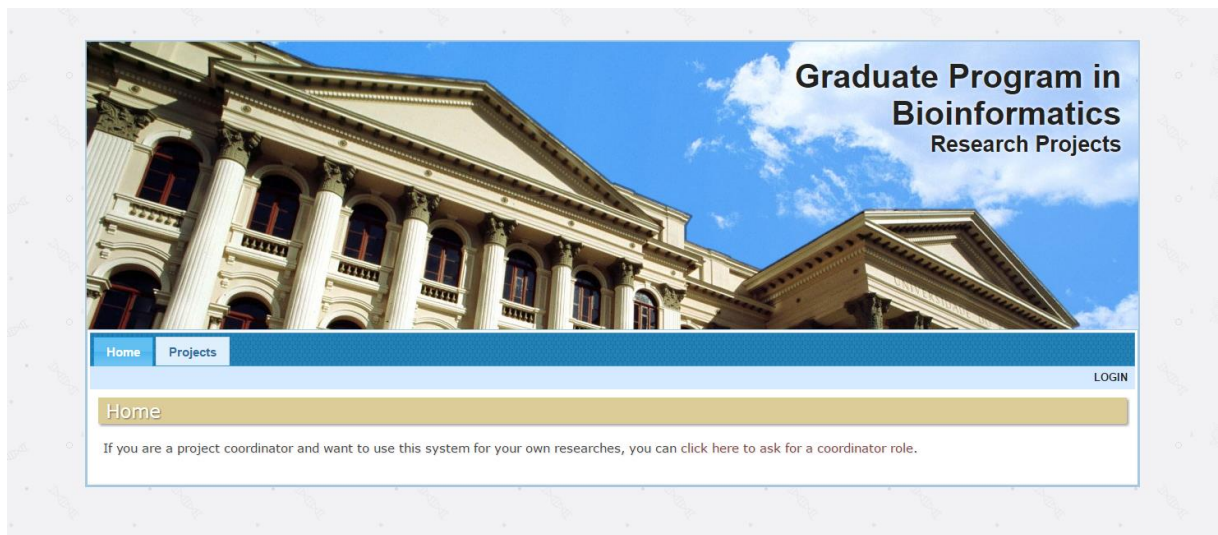


FIGURA 1 - TELA INICIAL PÚBLICO

Fonte: Os autores (2016).

Ao clicar em “Projects”, temos a tela da FIGURA 02. Apesar dessa tela estar disponível para todos os usuários, os projetos podem possuir restrição de visualização de acordo com o tipo do usuário.

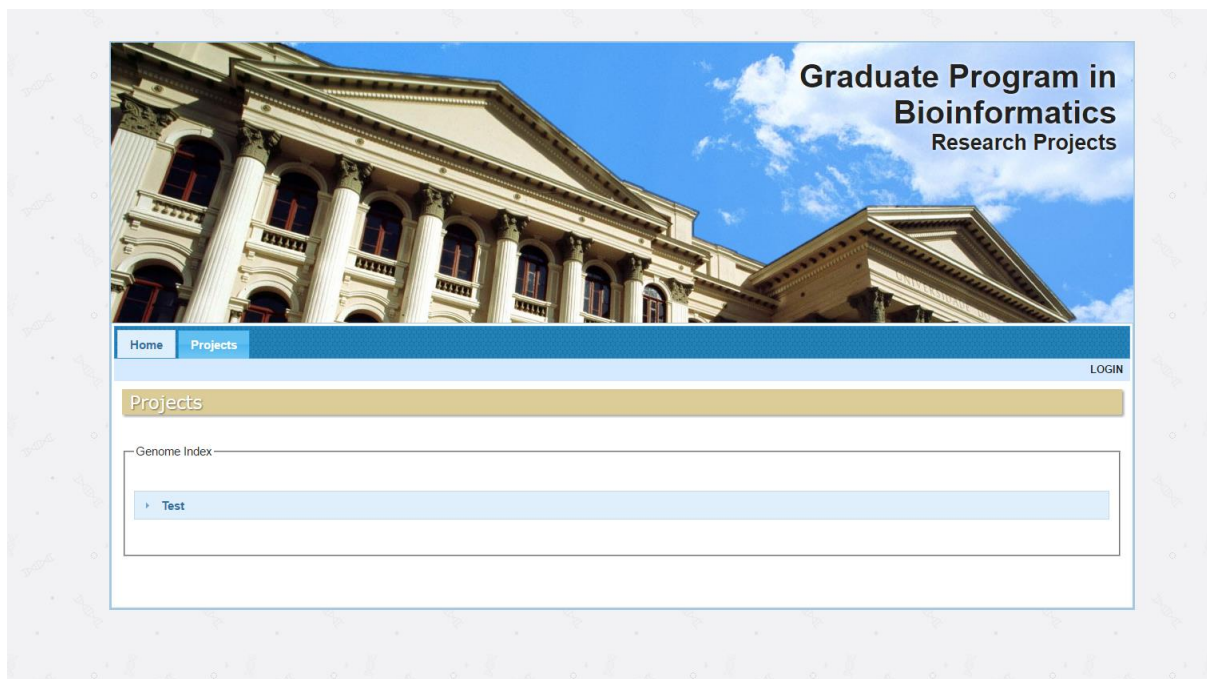


FIGURA 2 – TELA PROJETOS

Fonte: Os autores (2016).

É possível baixar os arquivos genéticos de acordo com as restrições de permissão.

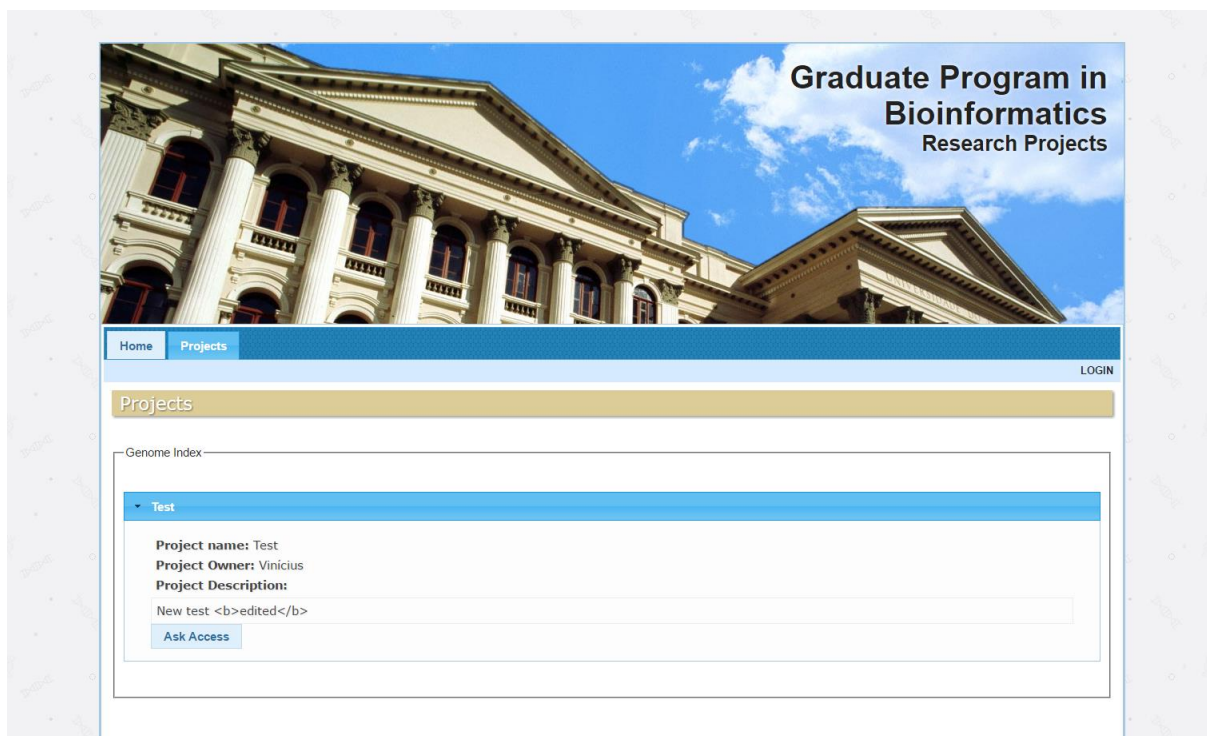
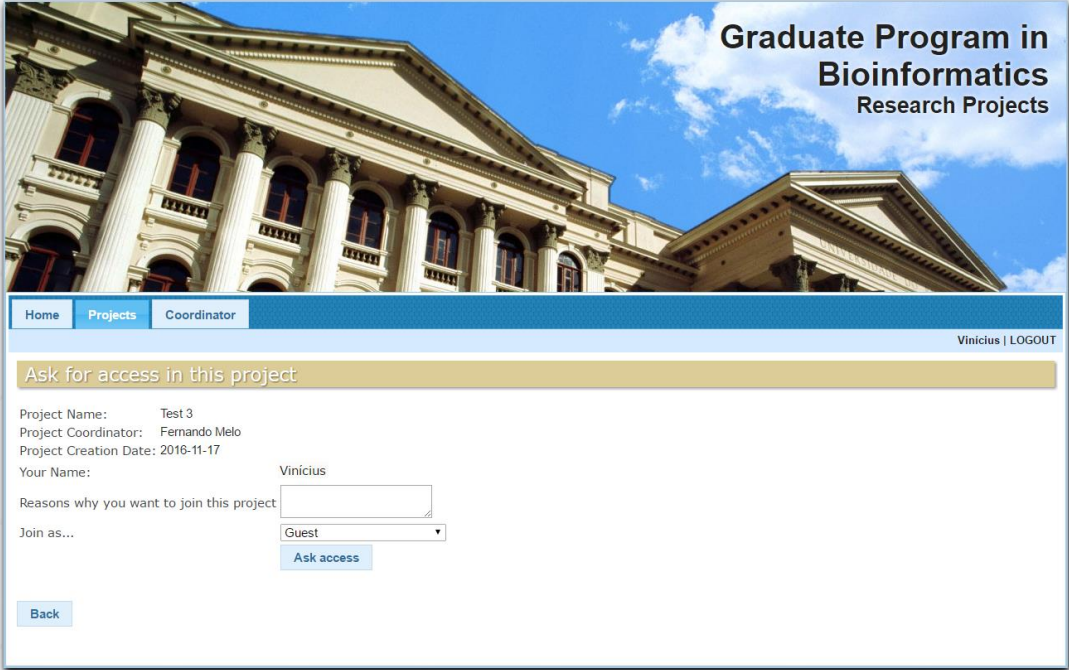


FIGURA 3 - TELA ARQUIVOS

Fonte: Os autores (2016).

E também é possível pedir acesso a um projeto. Caso o usuário esteja logado, é pedido um acesso de convidado (pois ele já é um pesquisador), caso o usuário não esteja logado é pedido um acesso de pesquisador e prossegue para a tela do FIGURA 5.



The screenshot displays a web application interface for the 'Graduate Program in Bioinformatics Research Projects'. The header features a background image of a classical building and the program's name. A navigation bar includes 'Home', 'Projects', and 'Coordinator' tabs. The user 'Vinicius' is logged in, with a 'LOGOUT' link. The main content area is titled 'Ask for access in this project' and contains the following information:

- Project Name: Test 3
- Project Coordinator: Fernando Melo
- Project Creation Date: 2016-11-17

Below this information, there is a form for requesting access:

- A text input field for 'Your Name:' with the value 'Vinicius'.
- A text area for 'Reasons why you want to join this project:'.
- A dropdown menu for 'Join as...' with 'Guest' selected.
- An 'Ask access' button.
- A 'Back' button at the bottom left.

FIGURA 4 - TELA SOLICITAR ACESSO DE CONVIDADO

Fonte: Os autores (2016).

É preenchido os dados do solicitante e mostrado quais são os dados do projeto solicitado.

Graduate Program in
Bioinformatics
Research Projects

Home Projects LOGIN

Ask for access in this project

Project Name: Test
Project Coordinator: Vinicius
Project Creation Date: 2016-11-17
Your Name:
Reasons why you want to join this project:
Join as...:

FIGURA 5 - TELA CADASTRO PESQUISADOR

Fonte: Os autores (2016).

Voltando a tela inicial, o visitante pode solicitar acesso como coordenador ao clicar em ao clicar em “click here to ask for a coordinator role”. Temos na tela da FIGURA 6 o cadastro para solicitação do acesso de coordenador para o administrador.

Graduate Program in
Bioinformatics
Research Projects

Home Projects LOGIN

Sign up as coordinator

Display name

E-mail

Password

Password confirmation

Create User

Back

FIGURA 6 - TELA SOLICITAR ACESSO DE COORDENADOR

Fonte: Os autores (2016).

5.2 ACESSO DE ADMINISTRADOR

Ao logar com o perfil de administrador, é apresentado a tela da FIGURA 7, mostrando as possibilidades de administração.

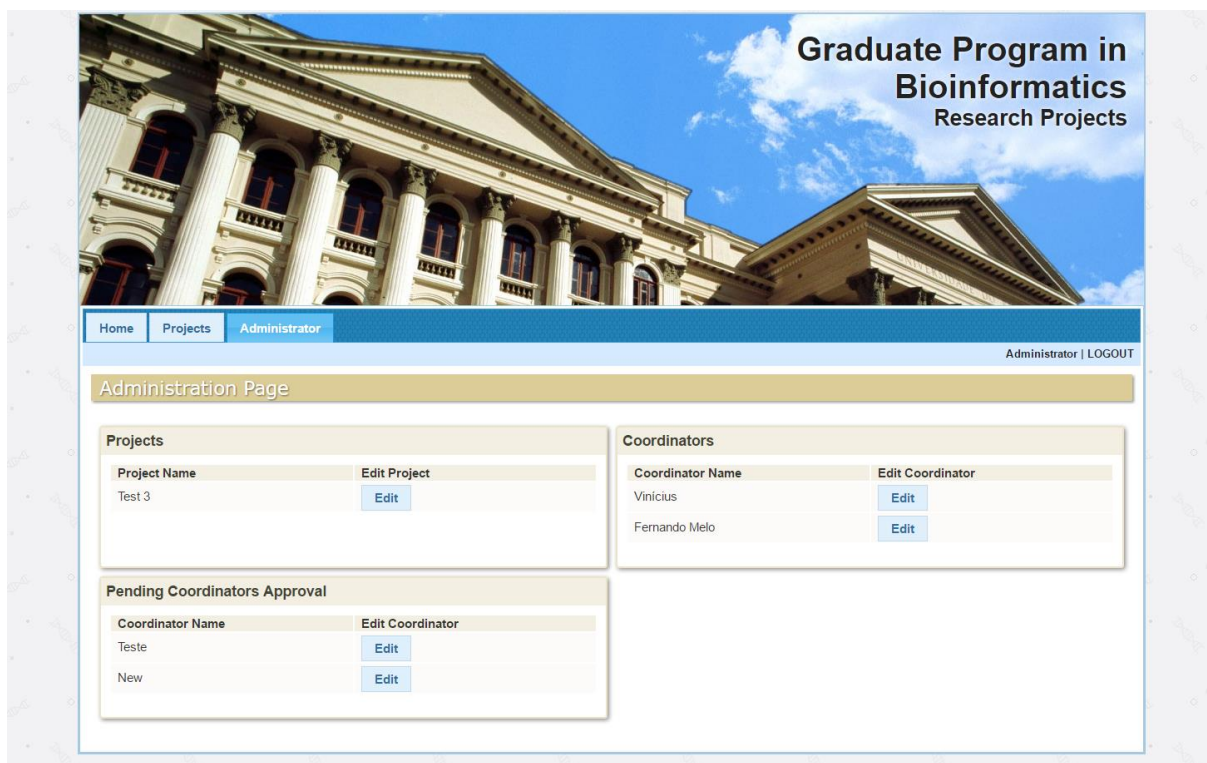


FIGURA 7 - TELA INICIAL ADMINISTRADOR

Fonte: Os autores (2016).

Caso escolha por editar um projeto, o administrador pode mudar o coordenador de um projeto.

Biomimimatics Research Projects

Home Projects Administrator Administrator | LOGOUT

Edit Project

Use the form below to view or modify project information.

Project details:

Project Name	Test 3
Description	
Start date	11/17/2016
Development Status	Ongoing
Current Project Leader	Fernando Melo

Save

Delete Project

Back

FIGURA 8 - TELA EDITAR PROJETOS

Fonte: Os autores (2016).

Na tela da FIGURA 9 o administrador pode aceitar novos coordenadores e novos projetos.

Biomimimatics Research Projects

Home Projects Administrator Administrator | LOGOUT

Approve or Reject Coordinator

Use the checkbox below to toggle user activity in the system. Check it to allow as a coordinator.
If you left unchecked, the user will not be able to log on to the system.
If you check the box, the user will be able to log on as a coordinator and you can find it under your "Coordinators" area, in the Administration pane

User details:

Name	Teste
E-mail	teste@ufpr.br
Reasons	
Active Coordinator?	<input type="checkbox"/> (check to enable)

Save

You can also delete this coordinator from the system as long as he/she has no projects assigned.
(You can delete this user. There is no projects assigned to him/her.)

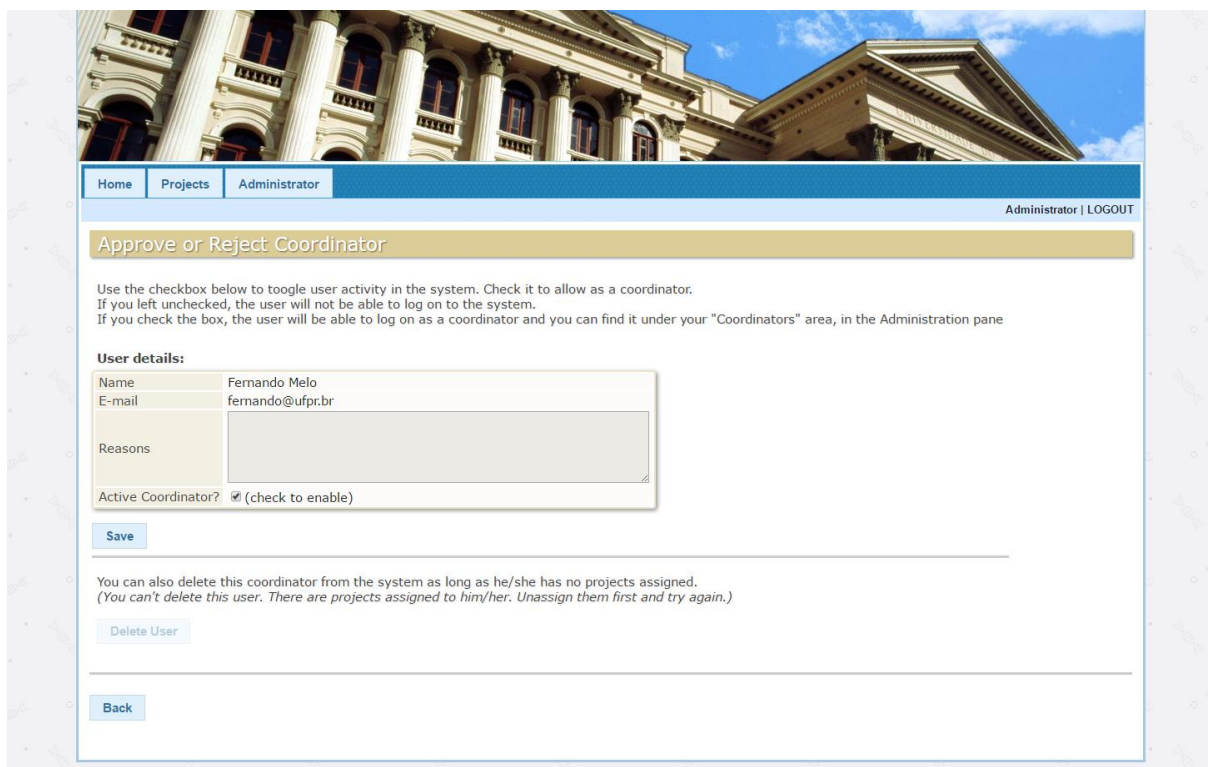
Delete User

Back

FIGURA 9 - TELA ACEITAR PROJETOS E NOVOS COORDENADORES

Fonte: Os autores (2016).

Caso escolha por editar ou excluir um coordenador, é possível ao administrador efetua essa ação na tela da FIGURA 10.



The screenshot displays a web application interface for managing coordinators. At the top, there is a navigation bar with tabs for 'Home', 'Projects', and 'Administrator'. The 'Administrator' tab is selected, and the user is logged in as 'Administrator | LOGOUT'. The main heading is 'Approve or Reject Coordinator'. Below this, a text block explains the functionality: 'Use the checkbox below to toggle user activity in the system. Check it to allow as a coordinator. If you left unchecked, the user will not be able to log on to the system. If you check the box, the user will be able to log on as a coordinator and you can find it under your "Coordinators" area, in the Administration pane'. The 'User details' section shows the name 'Fernando Melo' and email 'fernando@ufpr.br'. There is a 'Reasons' text area and an 'Active Coordinator?' checkbox which is checked, with the label '(check to enable)'. A 'Save' button is located below the details. A warning message states: 'You can also delete this coordinator from the system as long as he/she has no projects assigned. (You can't delete this user. There are projects assigned to him/her. Unassign them first and try again.)'. Below this is a 'Delete User' button. At the bottom, there is a 'Back' button.

User details:	
Name	Fernando Melo
E-mail	fernando@ufpr.br
Reasons	
Active Coordinator?	<input checked="" type="checkbox"/> (check to enable)

[Save](#)

You can also delete this coordinator from the system as long as he/she has no projects assigned.
(You can't delete this user. There are projects assigned to him/her. Unassign them first and try again.)

[Delete User](#)

[Back](#)

FIGURA 10 - TELA EDITAR/ EXCLUIR COORDENADORES

Fonte: Os autores (2016).

5.3 ACESSO DE COORDENADOR

Depois de logado, é apresentado ao coordenador a tela da FIGURA 11, mostrando as possibilidades de administração.

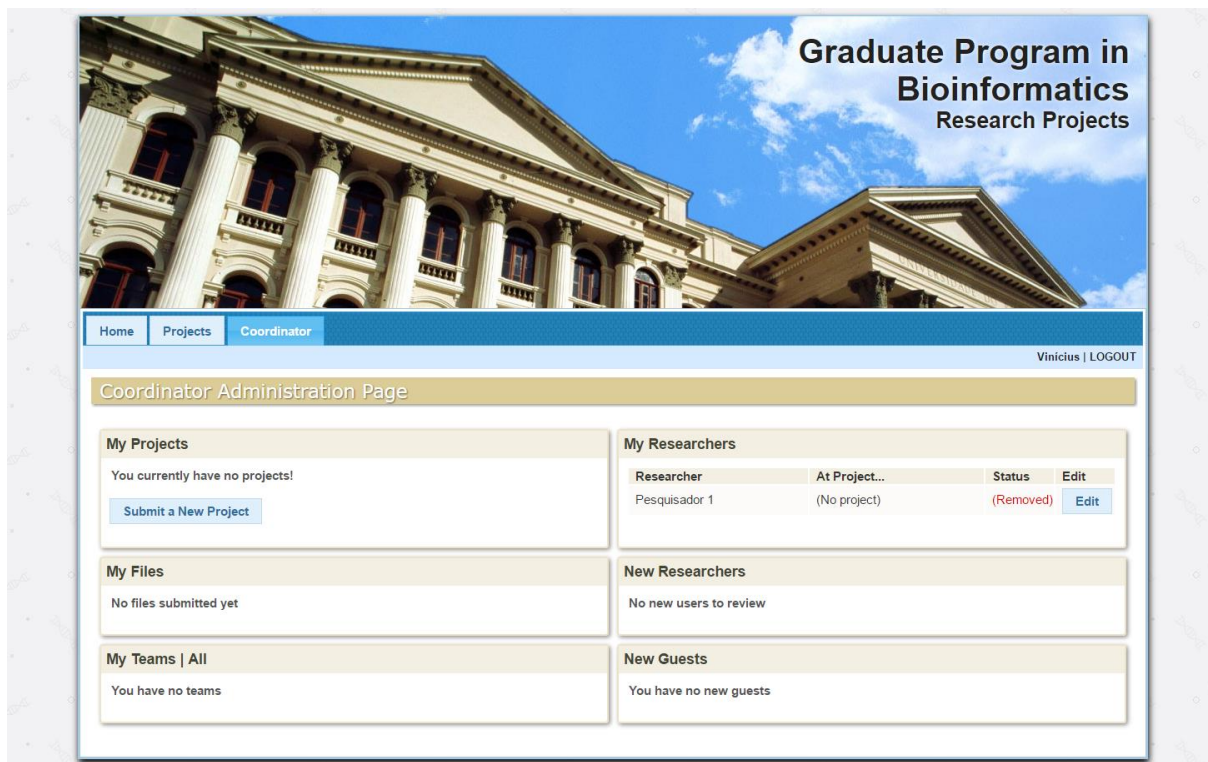


FIGURA 11 - TELA

Fonte: Os autores (2016).

Ao clicar em editar um projeto é possível ao coordenador editar as informações de um projeto ou excluí-lo.

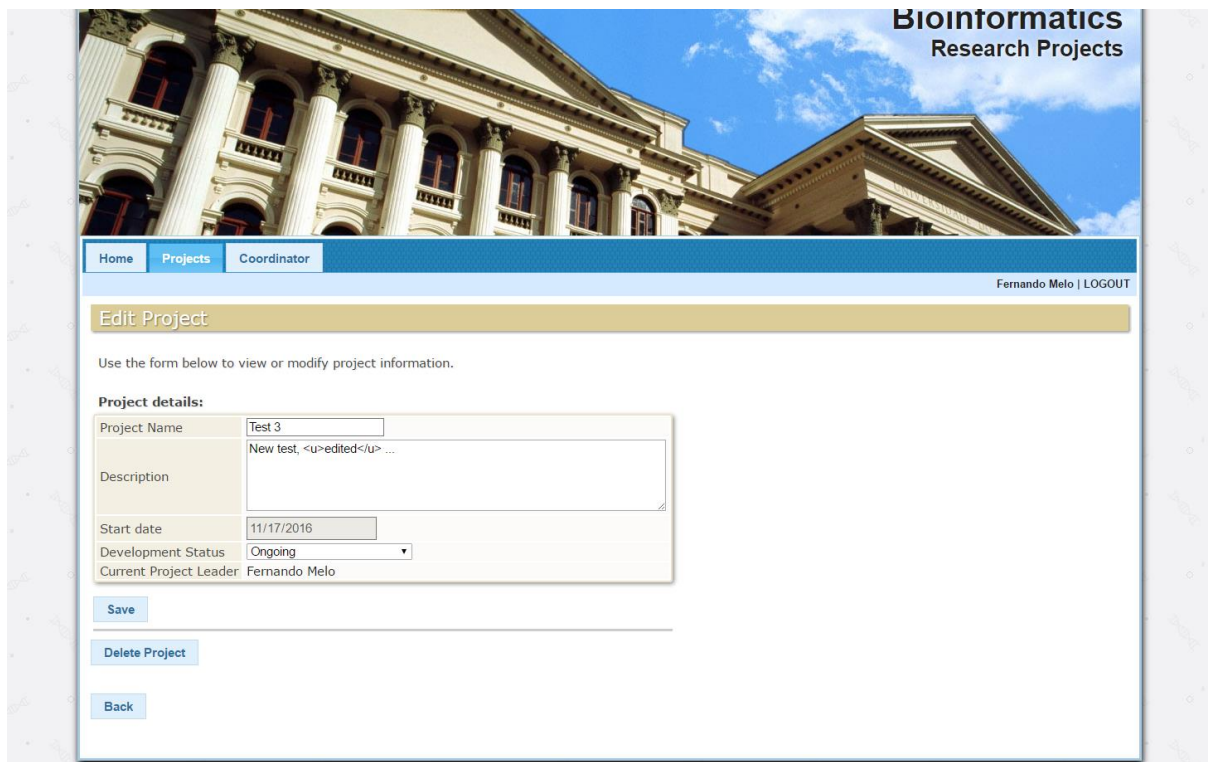


FIGURA 12 - TELA EDITAR/EXCLUIR PROJETOS

Fonte: Os autores (2016).

Ao clicar em editar um organismo é possível ao coordenador editar as informações de um organismo ou exclui-lo.

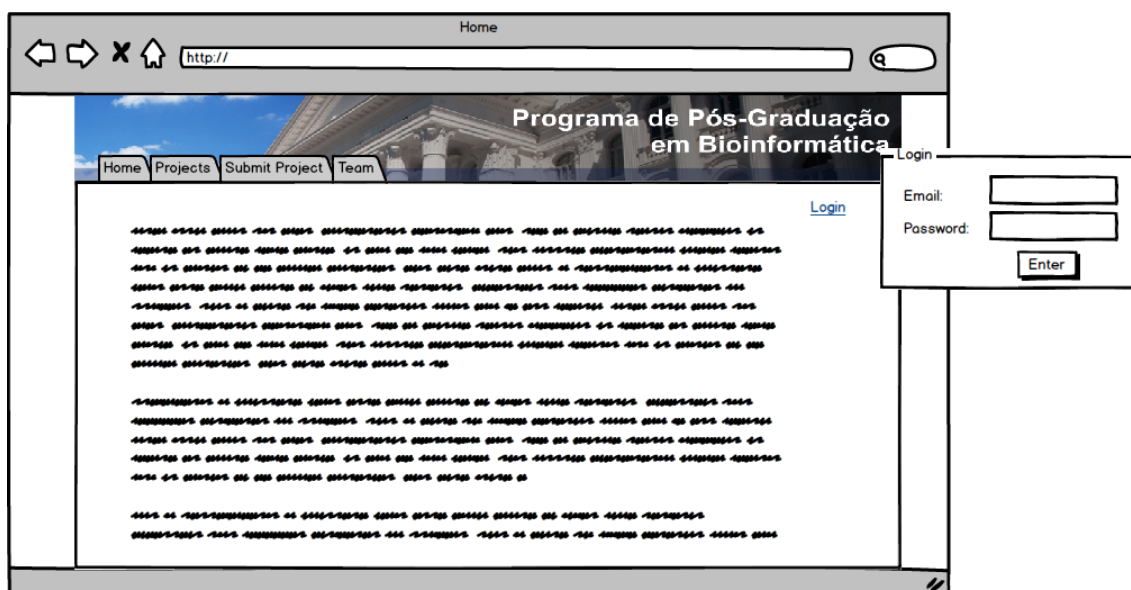
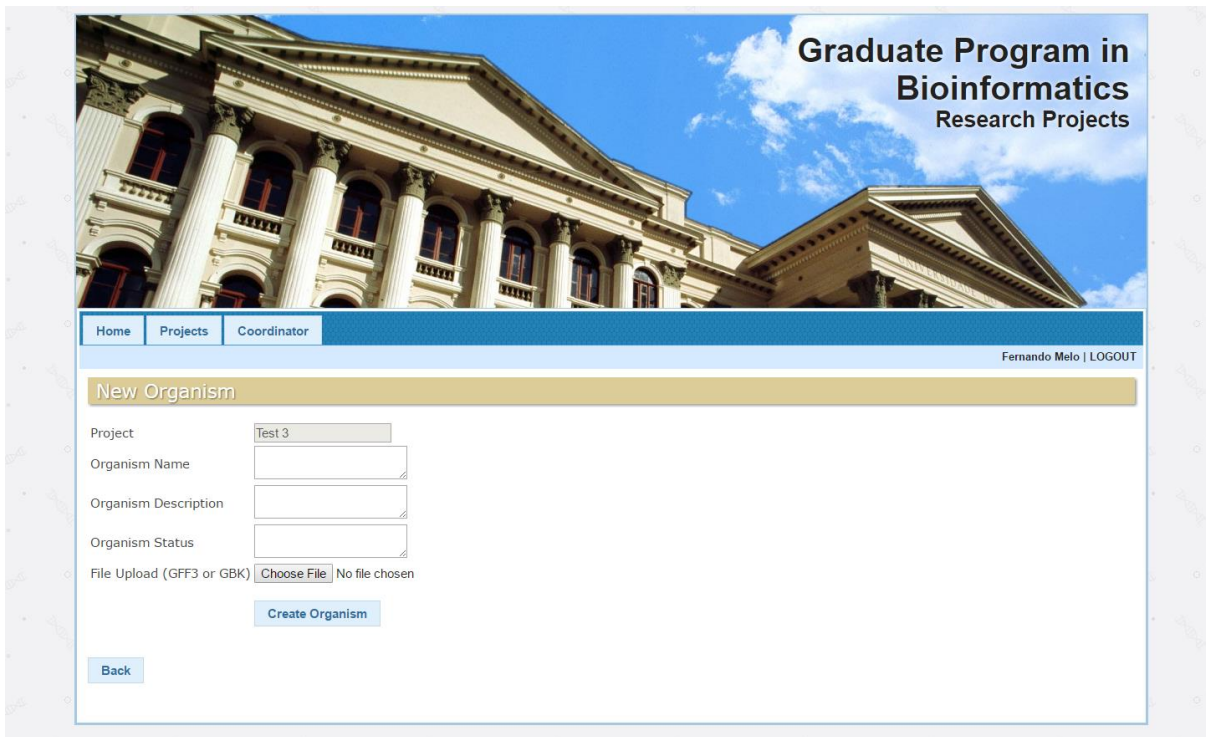


FIGURA 13 - TELA EDITAR/EXCLUIR ORGANISMOS

Fonte: Os autores (2016).

Enquanto edita um organismo o coordenador pode incluir um novo arquivo para aquele organismo, como mostrado na tela da FIGURA 14.



The screenshot displays a web application interface for the 'Graduate Program in Bioinformatics Research Projects'. The header features a background image of a classical building and the program's name. A navigation bar includes 'Home', 'Projects', and 'Coordinator' tabs, with 'Coordinator' being the active tab. The user 'Fernando Melo' is logged in, with a 'LOGOUT' link. The main content area is titled 'New Organism' and contains a form with the following fields: 'Project' (pre-filled with 'Test 3'), 'Organism Name', 'Organism Description', and 'Organism Status'. Below these is a 'File Upload (GFF3 or GBK)' section with a 'Choose File' button and the text 'No file chosen'. At the bottom of the form are 'Create Organism' and 'Back' buttons.

FIGURA 14 - TELA INCLUIR ARQUIVOS DE PROJETO

Fonte: Os autores (2016).

Após submeter um novo projeto (FIGURA 15) o coordenador monta a equipe para aquele projeto, como mostrado na tela da FIGURA 16

Graduate Program in Bioinformatics Research Projects

Home Projects Coordinator

Vinicius | LOGOUT

Create a New Project

Project name

Description

Start date

Development status

FIGURA 15 - TELA SUBMETER UM NOVO PROJETO

Fonte: Os autores (2016).

Programa de Pós-Graduação em Bioinformática

Home Projects Submit Project Team

Login

Email:

Password:

FIGURA 16 - TELA CADASTRA EQUIPE

Fonte: Os autores (2016).

Ao receber uma solicitação de acesso de pesquisador/convidado a um de seus projetos, o coordenador pode decidir se aceita ou não.

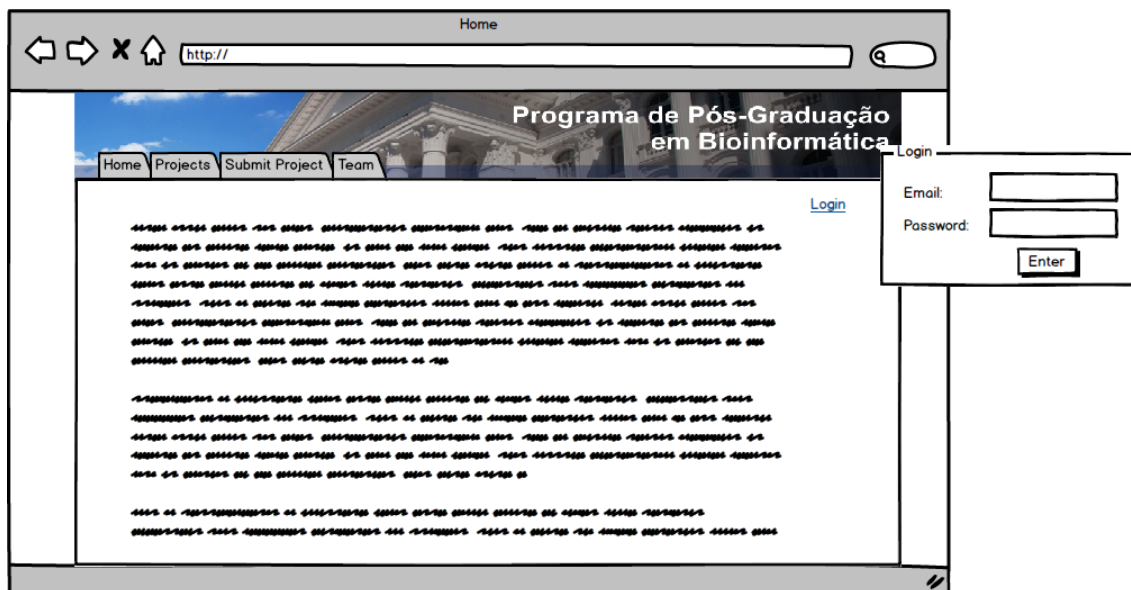


FIGURA 17 - TELA ACEITAR PESQUISADORES/CONVIDADOS

Fonte: Os autores (2016).

O coordenador também pode excluir seus pesquisadores, como visto na tela da FIGURA 18.

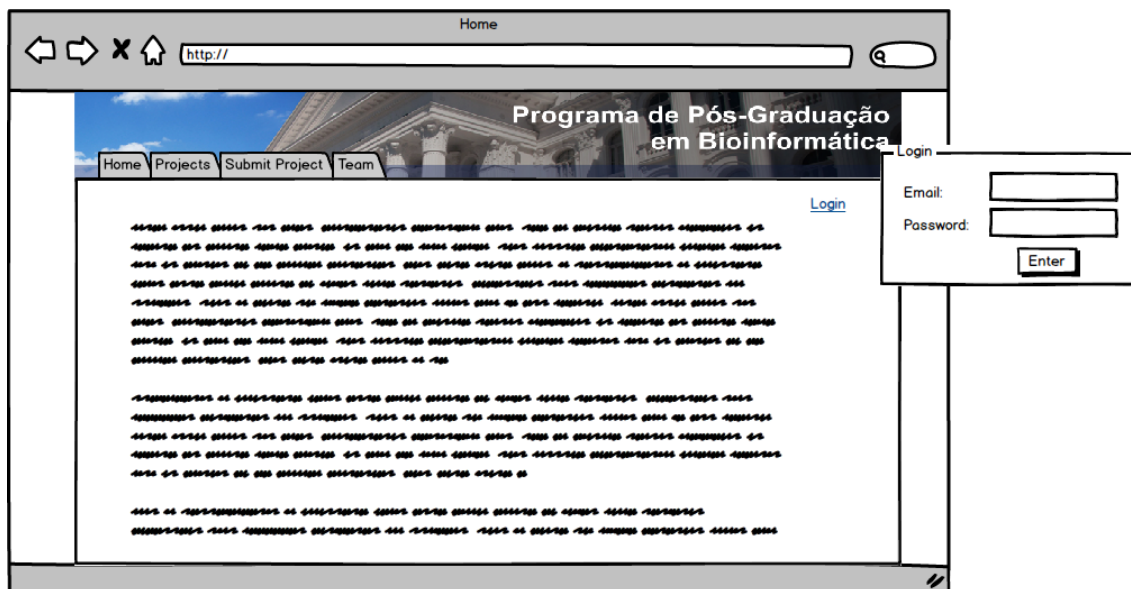


FIGURA 18 – TELA EXCLUIR PESQUISADORES

Fonte: Os autores (2016).

O coordenador também consegue editar as equipes de seus projetos

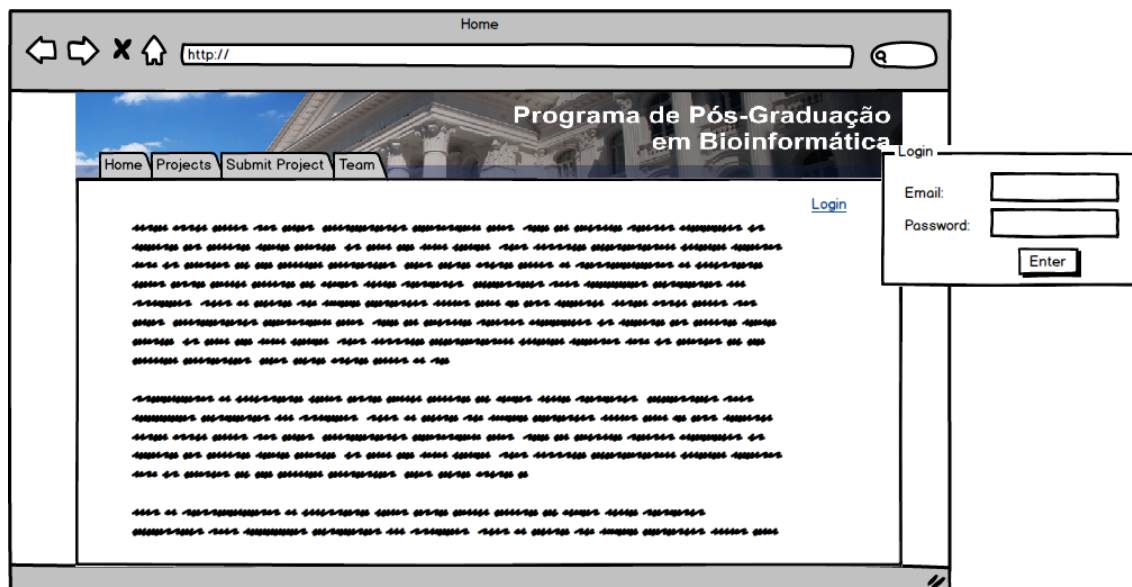


FIGURA 19 - TELA EDITAR EQUIPES

Fonte: Os autores (2016).

5.4 ACESSO DE PESQUISADOR

No momento as atividades possíveis a um pesquisador não variam muito comparado a de um usuário público, sendo a única diferença no momento a possibilidade de ver mais projetos devido a sua restrição.

5.5 INSTALAÇÃO

A instalação foi realizada no ambiente da PPG-BIOINFO rodando em servidor Linux, para instalação em outros tipos de plataformas configurações diferenciadas serão necessárias. O que estiver entre <> é variável e deve ser substituído por valores desejados:

Instalação do Ruby e da gem do Rails:

- `sudo apt-get install ruby2.3 ruby2.3-dev`
- `gem install rails`

Instalação do Javascript Runtime NodeJS:

- `sudo apt-get install nodejs`

Instalação do banco de dados PostgreSQL:

- `sudo apt-get install aptitude`
- `sudo apt-get install postgresql postgresql-contrib`
- `sudo apt-get install libpq-dev` (Para conseguir compilar a gem "pg" do Postgres)
- `sudo gem install pg`

Configuração do banco de dados PostgreSQL:

- `sudo -u postgres createuser -s <aluno>`
 - (cria o usuário "aluno", usar o mesmo da conta do sistema ou criar outro usuario no linux)
- `sudo -u postgres psql`
 - (entra no console do postgres, digitar os comandos abaixo para criar uma senha e sair do console do Postgres)
 - `\password <aluno>`
 - `<inserir uma senha>`
 - `<confirmar>`
 - `\q`

Configuração da aplicação:

Deve-se navegar até a pasta raiz da aplicação, ou seja base_fungos, e alterar o arquivo do banco de dados com usuário e senha criados:

- `sudo vim config/database.yml`

Alterar as linhas abaixo com os valores de usuario e senha, prestando atenção para não usar tabulação por 'tab' no arquivo, e sim utilizar o espaçamento correto, pois se trata de um formato de serialização YAML:

- `username: <aluno>`
- `password: <senha do <aluno> no PostgreSQL>`

No diretorio da aplicação (base_fungos):

- `bundle install` (esse passo irá ler a configuração da aplicação e instalar as gems necessárias)

Instalar o banco de dados da aplicação no PostgreSQL:

- `rake db:create`
- `rake db:migrate`
- `rake db:seed`

Depois de instalar e executar o `rake db:seed`, alguns dados serão inseridos nas tabelas:

- Valores iniciais para as tabelas de:
 - `user_roles`
 - `team_roles`
 - `organism_statuses`
- Usuário de administrador:
 - `username: admin`
 - `password: senha`

Estará tudo configurado e pronto. Para iniciar a aplicação, basta dar o comando abaixo, ainda na raiz da aplicação:

- `rails server` (ou `rails server -b xxx.xxx.xxx.xxx` para executar o servidor no IP do host atual permitindo acesso externo)

- Para parar o servidor, basta apertar CTRL+C no console

Depois de iniciada a aplicação, basta acessá-la:

- Para acesso interno:
 - localhost:3000
 - 127.0.0.1:3000
- Para acesso externo:
 - 192.168.2.23:3000

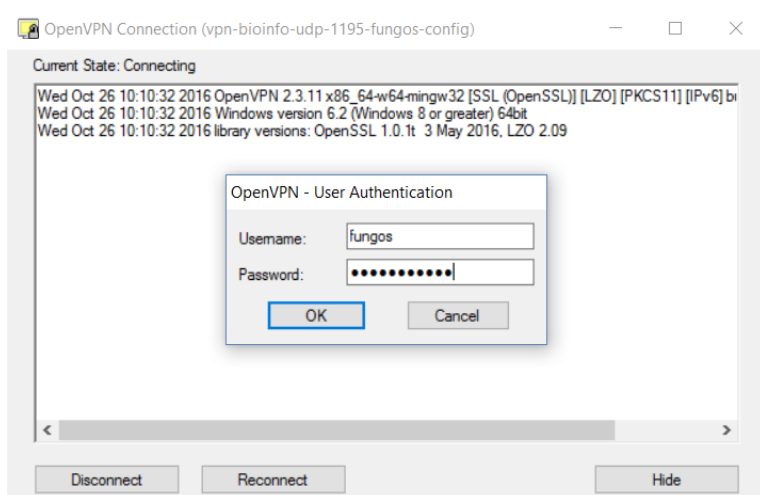
5.6 ACESSO

Para acessar a aplicação é necessário estar na rede interna da BIO-INFO. Caso não tenha acesso a rede Wi-Fi “UFPR_SEPT_BIOINFOLAB” é possível se conectar através de uma VPN. Para todos os acessos utilizamos a OpenVPN. No link abaixo o arquivo já vem pré-configurado com as informações de acesso necessária:

- <https://200.236.3.122/owncloud/index.php/s/NiJciRN7GkCPaQd>

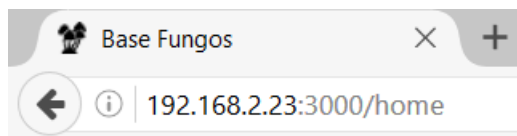
A instalação é padrão, não é necessária nenhuma configuração durante a instalação. Mantenha o ícone da OpenVPN no desktop e garanta que a VPN seja executada como administrador. O usuário e senha para acesso é:

- Usuário: fungos
- Senha: FungBio#@20



Depois de estar na rede, via VPN ou de forma local, acessar o endereço:

- <http://192.168.2.23:3000/>



5.6.1 Acessos de Usuário:

Os seguintes usuários já estão pré-configurados para acesso ao sistema:

- Administrador:
 - Usuário: admin@ufpr.br
 - Senha: senha
- Coordenador:

- Usuário: fernando@ufpr.br
 - Senha: senha
- Pesquisador:

6 CONSIDERAÇÕES FINAIS

O desenvolvimento do sistema nos possibilitou ter um contato inicial com a área da bioinformática e perceber a importância dessa tanto para o desenvolvimento das mais diversas áreas da biologia como para a criação de novas tecnologias.

O aplicativo consegue cumprir as funções de armazenamento de informações genéticas dos organismos pesquisados assim como auxiliar no gerenciamentos de times e projetos.

A conversão dos arquivos GBK para os formatos GFF e Fasta é muito bem vinda para os pesquisadores que precisam utilizar diferentes formatos de arquivos, principalmente para quem precisa somente utilizar a sequência genética de um organismo, e pode optar por fazer o download no formato Fasta. Porém, a impossibilidade de conversão entre os tipos de arquivo pode vir a parecer que a funcionalidade não está completa.

As funcionalidades de download e upload de arquivos genéticos consomem um alto valor computacional e de rede, por isso o sistema pode vir a enfrentar problemas de lentidão em caso de muitos acessos simultâneos a essas funções, mesmo com a implementação da compactação.

Durante o desenvolvimento, várias ideias de como o software poderia ser melhorado e ampliado surgiram, porém não foi possível implementá-las devido a complexidade e tempo. Algumas destas ideias para uma futura versão estão listadas abaixo:

- Criar uma ferramenta para busca por genomas ou descrição de genes
- Utilizar um banco de dados non-SQL afim de diminuir o impacto computacional para Download e Upload
- Implementar a conversão de GFF para GBK

REFERÊNCIAS

ASTAH. **Astah – Fundamentals**. Disponível em: <<http://astah.net/fundamentals>>. Último acesso em: 27/06/2016.

BALSAMIQ. **Balsamiq Mockups**. Disponível em: <<https://balsamiq.com/products/mockups/>>. Último acesso em: 26/06/2016.

PPG-BIOINFO. Programa de Pós-Graduação em Bioinformática. Disponível em: <<http://www.bioinfo.ufpr.br/about.php>>. Último acesso em: 26/06/2016.

BIORUBY. **BioRuby: bioinformatics software for the Ruby programming language**. Disponível em: <<http://bioruby.org/>>. Último acesso em: 20/06/2016.

BROAD INSTITUTE. **Black Yeast Database**. Disponível em: <http://www.broadinstitute.org/annotation/genome/Black_Yeasts/MultiHome.html>. Último acesso em: 26/06/2016.

GIT. **About – Git**. Disponível em: <<http://git-scm.com/about/>>. Último acesso: 20/05/2016.

GUEDES, G. **UML 2: uma abordagem prática**. São Paulo: Novatec, 2011.

IBM. **IBM – Definição de RUP**. Disponível em: <<http://www-01.ibm.com/software/rational/rup/>>. Último acesso: 21/05/2016.

MYSQL. **MySQL :: MySQL Workbench**. Disponível em: <<https://www.mysql.com/products/workbench/>>. Último acesso em: 25/06/2016.

ORACLE. **Oracle VM VirtualBox**. Disponível em: <<https://www.virtualbox.org/>>. Último acesso em: 24/06/2016.

POSTGRESQL. **PostgreSQL: About**. Disponível em: <<https://www.postgresql.org/about/>>. Último acesso em: 23/06/2016.

RASHIDI, H.; BUEHLER, L. **Bioinformatic Basics – Applications in Biological Science and Medicine**. Londres: CRC Press, 2000.

RUBY. **About Ruby**. Disponível em: <<https://www.ruby-lang.org/en/about/>>. Último acesso em: 20/06/2016.

RUBY ON RAILS. **Ruby on Rails**. Disponível em: <<http://rubyonrails.org/>>. Último acesso em: 20/06/2016.

Astah Professional

Disponível em: < <http://astah.net/download>>. Último acesso em: 27/06/2016.

Balsamiq Mockups

Disponível em: < <https://balsamiq.com/download/>>. Último acesso em: 26/06/2016.

GitHub

Disponível em: < <https://github.com/join>>. Último acesso: 20/05/2016.

MySQL Workbench

Disponível em: < <https://www.mysql.com/downloads/>>. Último acesso em: 25/06/2016.

Oracle VM VirtualBox

Disponível em: < <https://www.virtualbox.org/wiki/Downloads>>. Último acesso em: 24/06/2016.

PostgreSQL

Disponível em: < <https://www.postgresql.org/download/>>. Último acesso em: 23/06/2016.

APÊNDICE A - DIAGRAMA DE CASOS DE USO

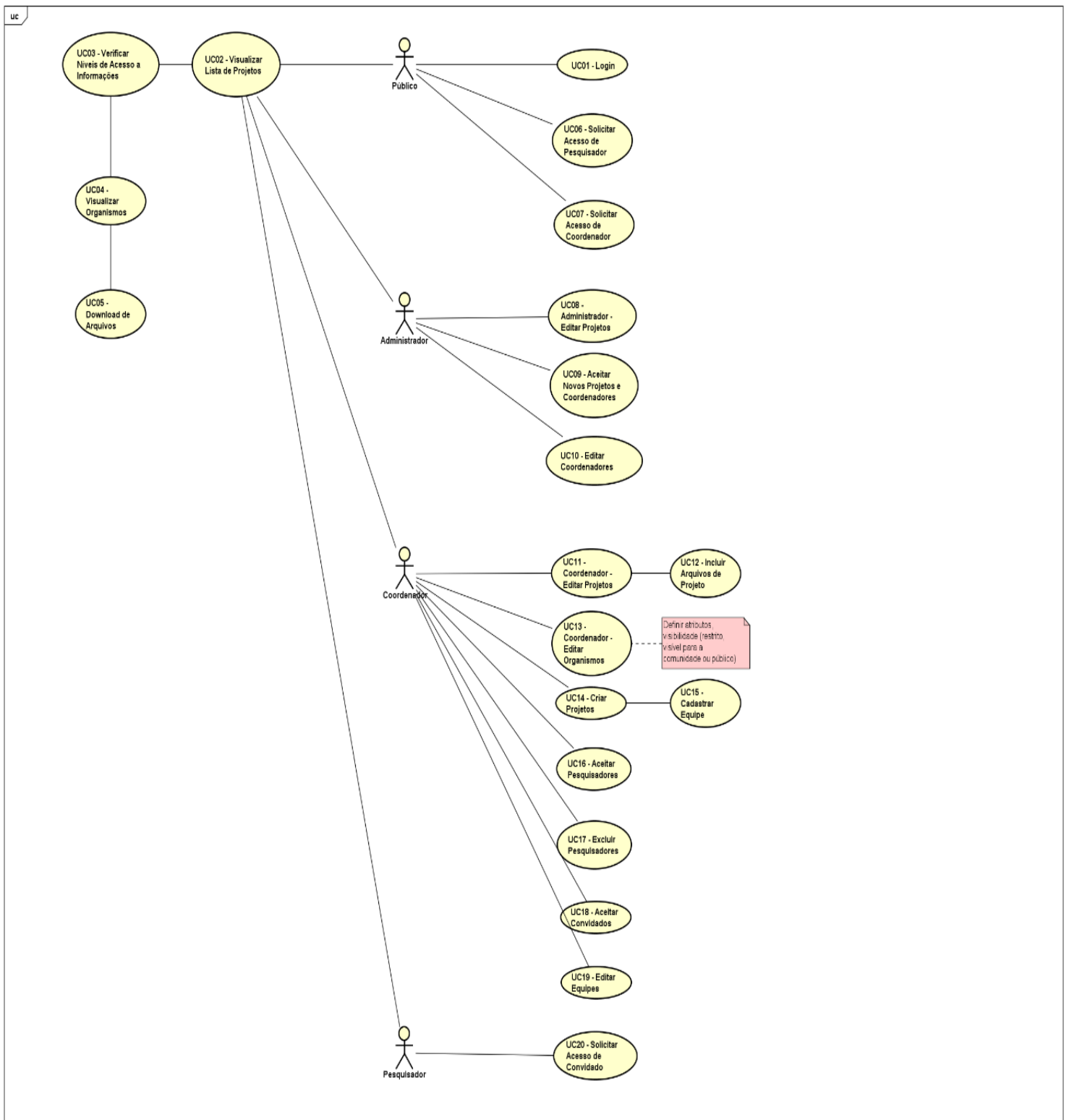


FIGURA xx – DIAGRAMA DE CASOS DE USO

Fonte: Os autores (2016).

APÊNDICE B - DIAGRAMA ENTIDADE RELACIONAMENTO

Valores iniciais das tabelas estarão no arquivo "db/seeds.rb" e serão inseridos durante a instalação através do comando:
"rake db:seed"

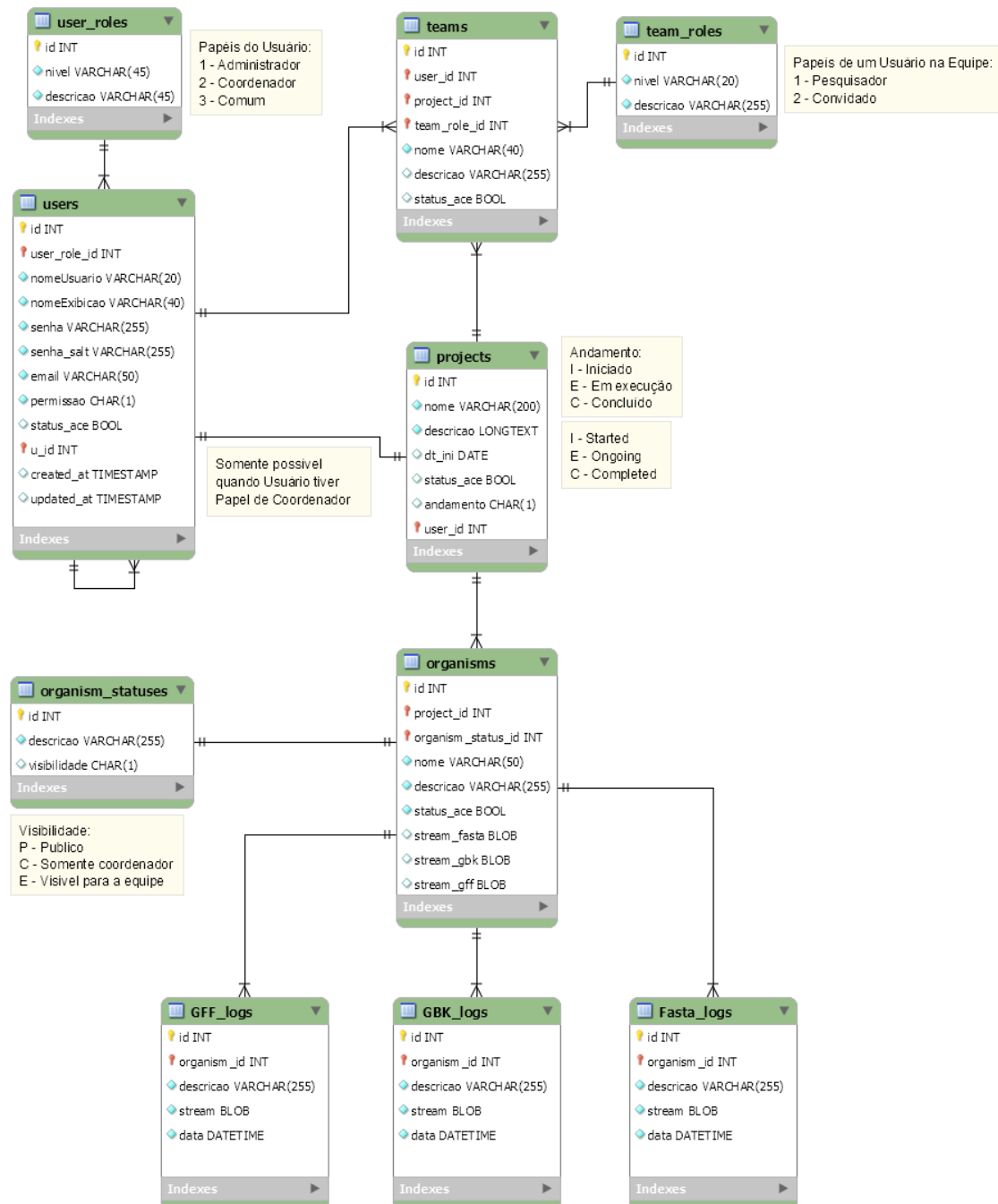


FIGURA xx – DIAGRAMA ENTIDADE RELACIONAMENTO

Fonte: Os autores (2016).

APÊNDICE C - DICIONÁRIO DE DADOS

TABELA 2 – DICIONÁRIO DE DADOS

TABELA	teams					
Faz a ligação entre os usuários comuns e um projeto						
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Time	Id	INT	PK			
ID Usuário	user_id	INT		(users/ id)		
ID Projeto	project_id	INT		(projects/ id)		
ID Papel	team_role_id	INT		(team_roles/ id)		
Nome	nome	VARCHAR(40)				
Obs do Usuário	obs_usr	VARCHAR (255)				
Status Acesso	status_ace	CHAR				

TABELA	team_roles					
Define o papel de cada usuário na equipe						
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Papel	id	INT	PK			
Nível	nivel	VARCHAR (20)				Pesquisador ou Convidado
Descrição	descricao	VARCHAR (255)				

TABELA	projects					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Projeto	id	INT	PK			
ID Coordenador	user_id	INT		(users/ id)	(Usuário/ papel) = "Coordenador"	
Nome	nome	VARCHAR (255)				
Descrição	descricao	VARCHAR (255)				
Data de Início	dt_ini	DATE				
Status de Aceitação	status_ace	CHAR				Ativo ou Inativo
Status de Andamento	status_and	CHAR				Em Andamento ou Finalizado

TABELA	users					
Caso o Usuário seja “Comum”, somente o Coordenador responsável pode exclui-lo. Caso o Usuário seja “Coordenador” somente um Administrador pode exclui-lo.						
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Usuário	id	INT	PK			
ID Coordenador	u_id	INT		(users / id)	(Usuário/ papel) <> “Coordenador” ou “Administrador”	Caso seja um Coordenador ou Admin, ID Coordenador será igual a null.
Papel	user_role_id	INT		(user_roles/ id)		
Nome de Usuário	nomeUsuario	VARCHAR (20)				
Nome de Exibição	nomeExibicao	VARCHAR (40)				
Email	email	VARCHAR (50)				*Login
Senha	senha	VARCHAR (40)				
Status de Aceitação	status_ace					Ativo ou Inativo

TABELA	user_roles					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Papel Usuário	id	INT	PK			
Nível	nivel	VARCHAR (45)				Adm, Coordenador ou Comum
Descrição	descricao	VARCHAR (45)				

TABELA	organisms					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Organismo	id	INT	PK			
ID Projeto	project_id	INT		(projects / id)		
ID Status	organism_status_id	INT		(organism_statuses / id)		
Nome	nome	VARCHAR (50)				
Descrição	descricao	VARCHAR (255)				
Status de Aceitação	status_ace					Ativou ou Inativo

ID Fasta	id_fasta	INT				
Stream Fasta	stream_fasta	BLOB				
Stream GBK	stream_gbk	BLOB				
Stream gff	stream_gff	BLOB				

TABELA	organism_statuses					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Status	id	INT	PK			
Descrição	descricao	VARCHAR (255)				
Visibilidade	Visibilidade	CHAR(1)				

TABELA	GFF_logs					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID GFF	id					
ID Organismo	organism_id			(organisms / id)		
Descrição	descricao					
Stream GFF	stream					
Data	date					

TABELA	GBK_logs					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID GBK	id					
ID Organismo	organism_id			(organisms / id)		
Descrição	descricao					
Stream GBK	stream					
Data	date					

TABELA	Fasta_logs					
C. Lógico	C. Físico	Tipo	PK	FK (Tabela/Campo)	Restrições	Observações
ID Fasta	id					
ID Organismo	organism_id			(organisms / id)		
Descrição	descricao					
Stream Fasta	stream					
Data	date					

APÊNDICE D - DIAGRAMA DE TELAS

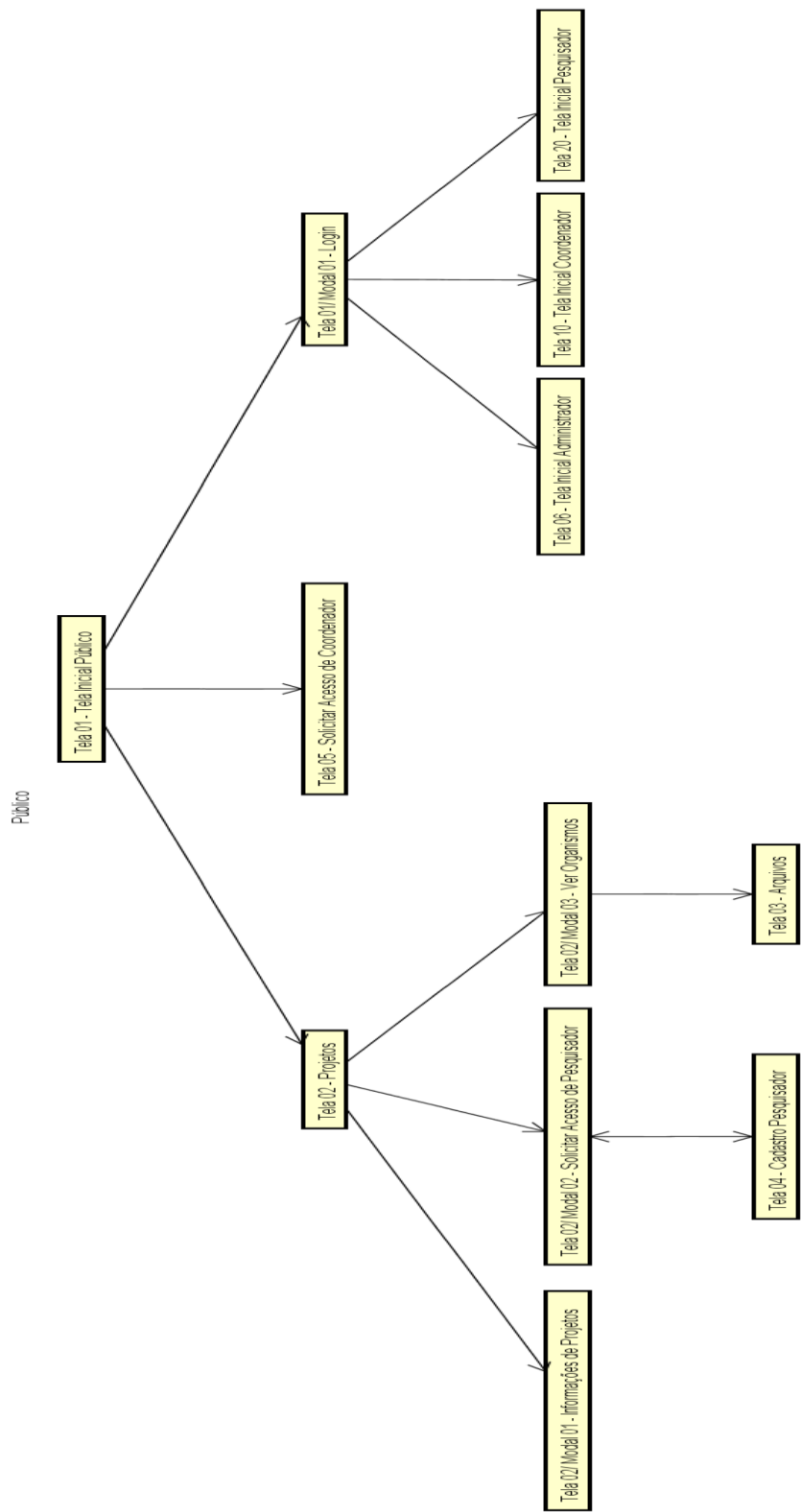


FIGURA xx – DIAGRAMA DE TELAS – ACESSO PÚBLICO

Fonte: Os autores (2016).

Administrador

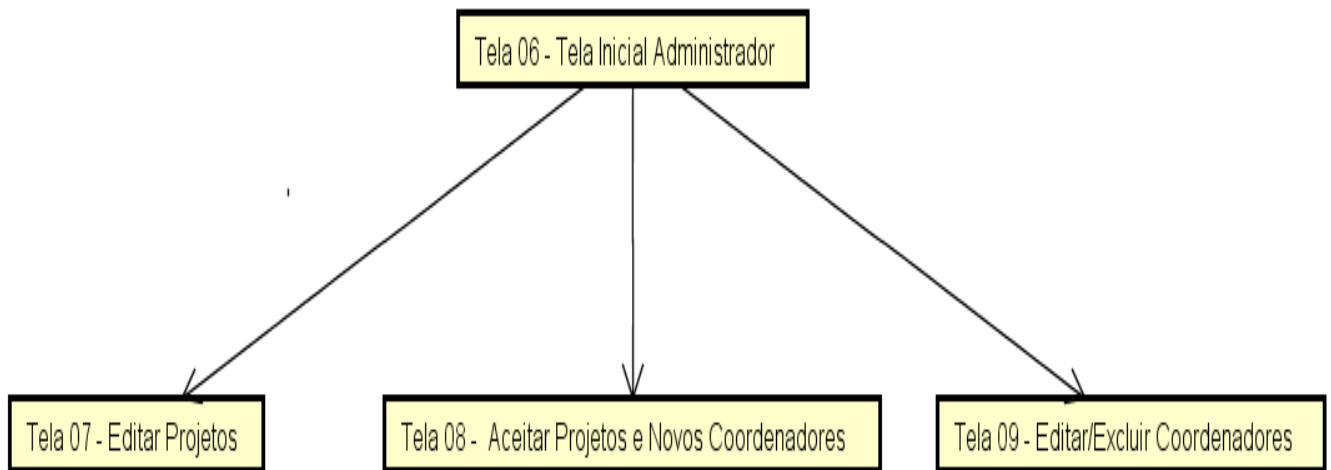


FIGURA xx – DIAGRAMA DE TELAS – ACESSO DE ADMINISTRADOR

Fonte: Os autores (2016).

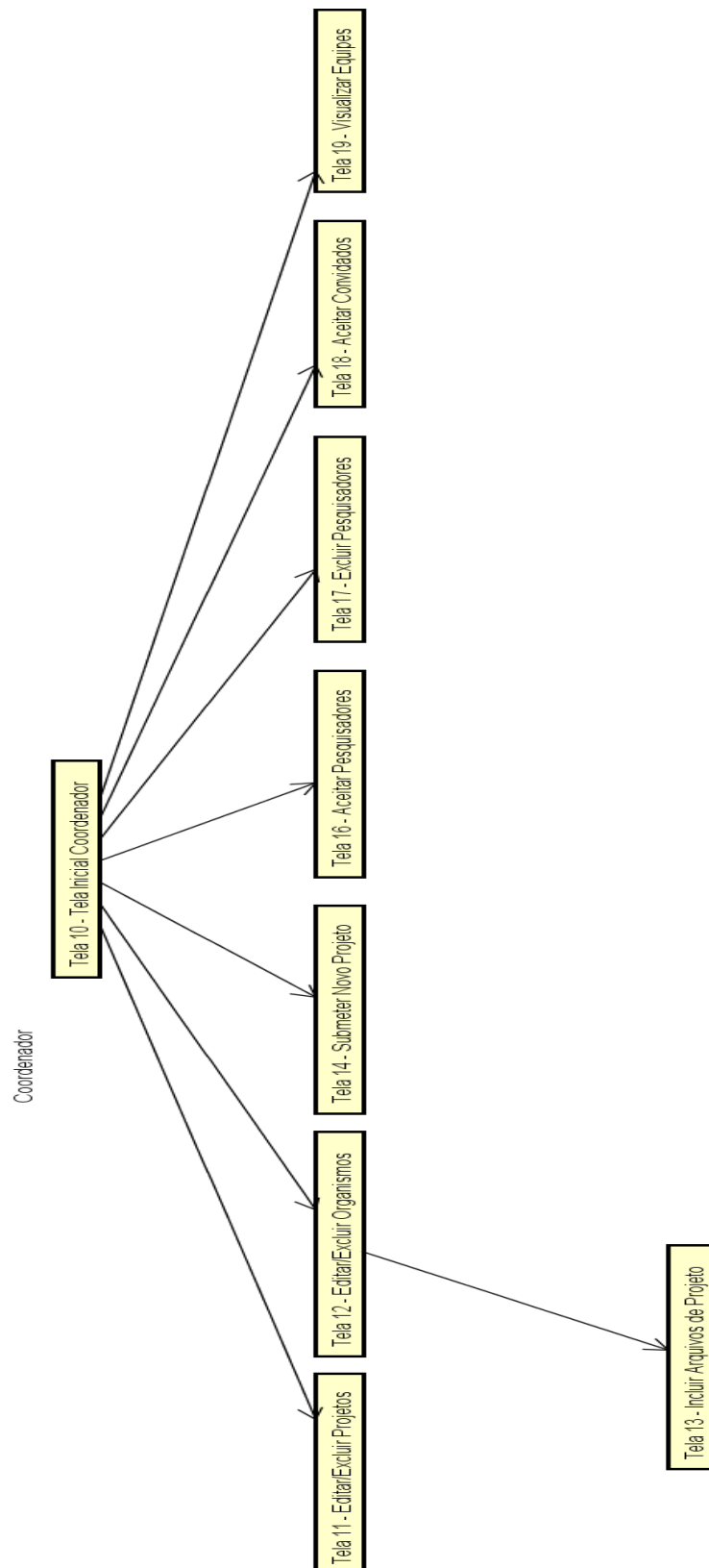


FIGURA xx – DIAGRAMA DE TELAS – ACESSO DE COORDENADOR

Fonte: Os autores (2016).

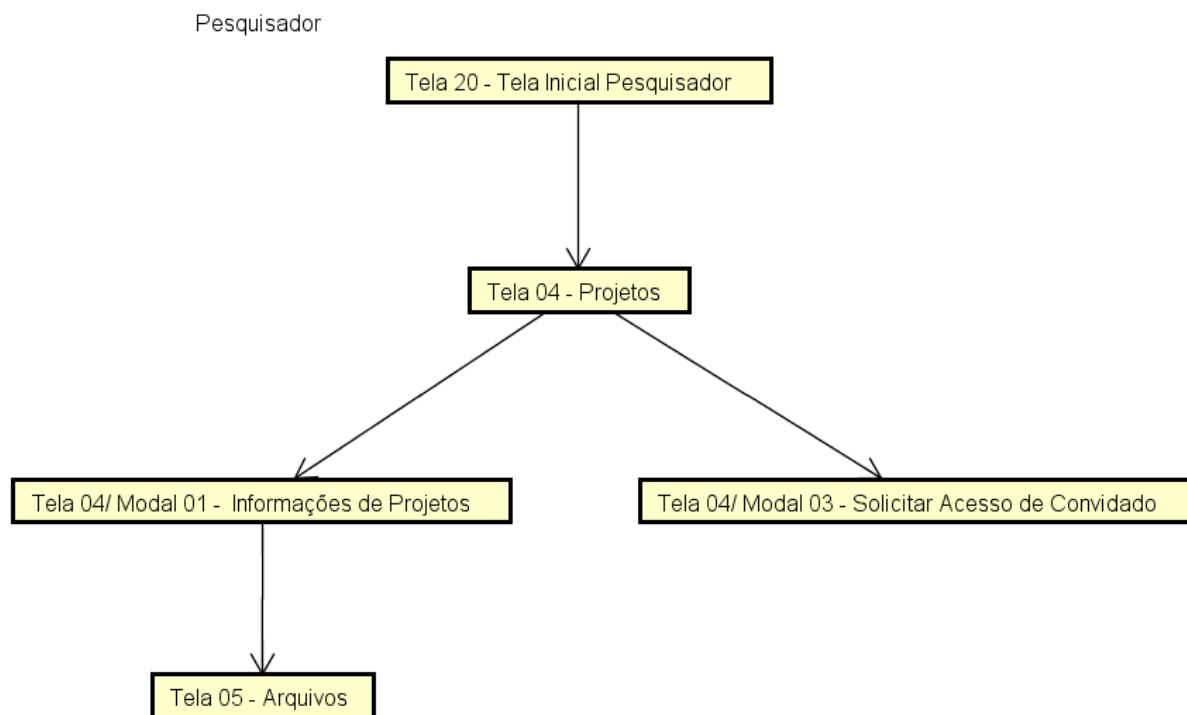


FIGURA xx – DIAGRAMA DE TELAS – ACESSO DE PESQUISADOR

Fonte: Os autores (2016).

APÊNDICE D - DIAGRAMA DE CLASSES

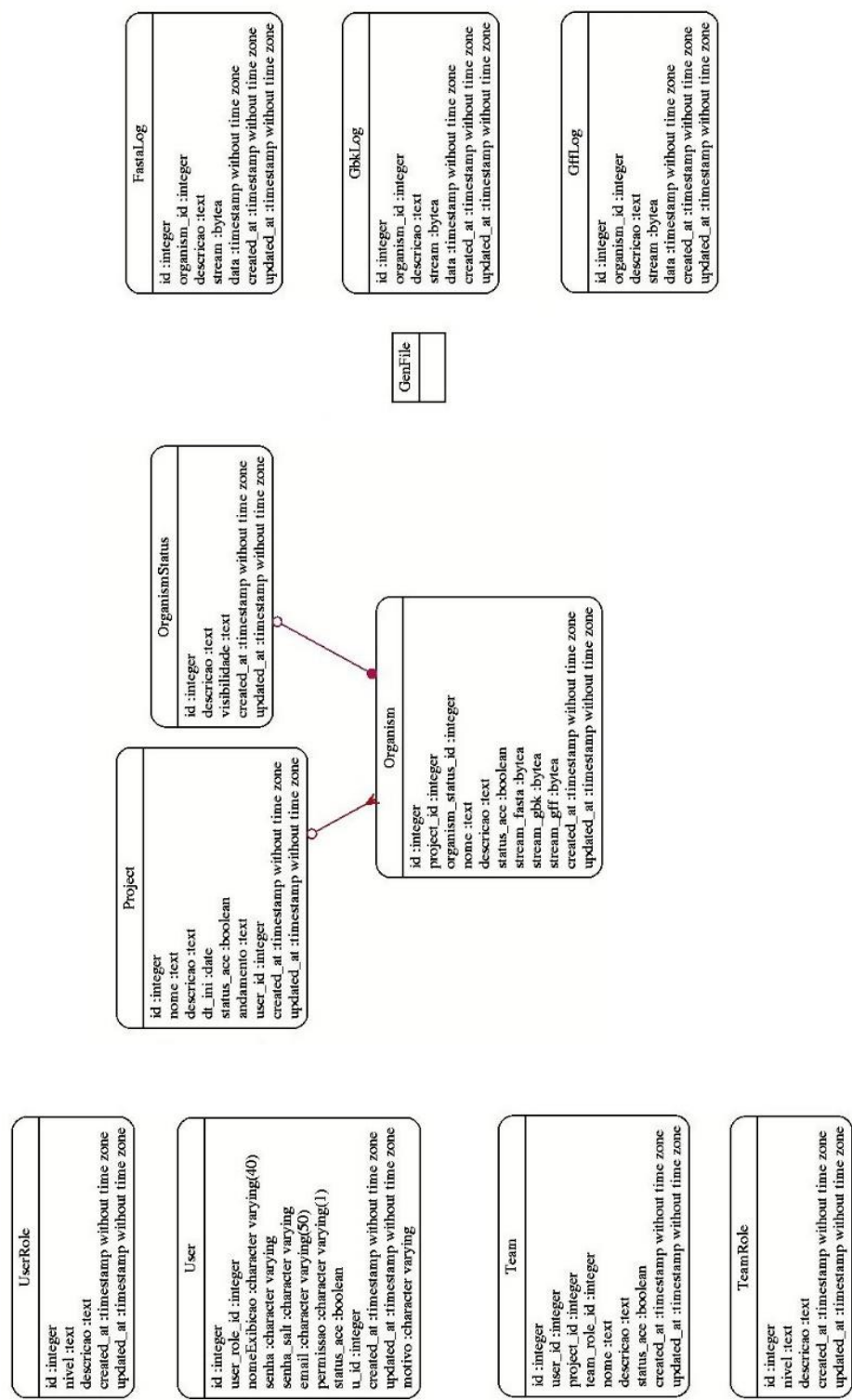


FIGURA xx – DIAGRAMA DE CLASSES

Fonte: Os autores (2016).