

UNIVERSIDADE FEDERAL DO PARANÁ

IRIS MEBS MORAES

PLATAFORMA PARA ARMAZENAMENTO E ANÁLISE DE DADOS GENÉTICOS

CURITIBA PR

2025

IRIS MEBS MORAES

PLATAFORMA PARA ARMAZENAMENTO E ANÁLISE DE DADOS GENÉTICOS

Trabalho de Conclusão de Curso apresentado à Universidade Federal do Paraná, como requisito para a obtenção de título de bacharel em Informática Biomédica .

Orientador: Prof. Dr. Lucas Ferrari de Oliveira.

Coorientador: Prof. Dr Ricardo Lehtonen Rodrigues de Souza.

CURITIBA PR

2025

AGRADECIMENTOS

Gostaria de agradecer aos professores envolvidos no projeto, Dr. Ricardo Souza e Dr. Lucas Ferrari. Ambos estiveram sempre à disposição quando foi necessário, oferecendo apoio e orientação.

Ao chegar ao fim de um dos ciclos mais importantes da minha vida, agradeço aos meus pais e às minhas avós, mesmo distantes, sempre me apoiaram e cuidaram de mim.

Aos meus amigos, que tornaram essa jornada mais leve, agradeço pela companhia, pelo incentivo constante e pelas memórias compartilhadas ao longo desses anos.

E ao meu companheiro André, que, além de namorado, divide a vida comigo. Ele esteve presente em todos os momentos, nos mais difíceis e nos mais bonitos.

RESUMO

A análise cromossômica e genômica tornou-se uma ferramenta fundamental no diagnóstico e tratamento personalizado na medicina clínica. Este trabalho propõe o desenvolvimento de uma plataforma para armazenamento, organização e análise de dados genéticos e fenotípicos, atendendo à demanda do Programa de Pós-Graduação em Genética da UFPR em parceria com a LAInfS. A motivação surge da necessidade de estruturar dados de pesquisas anteriores armazenados em planilhas, garantindo maior integridade, padronização e segurança. Para isso, foi modelado um banco de dados relacional com base nas planilhas fornecidas por docentes e pesquisadores, incorporando funcionalidades como importação e exportação de dados em formato CSV, gerenciamento de usuários e controle de acesso. A plataforma foi desenvolvida em Python, utilizando o framework Django. O trabalho também apresenta uma revisão sobre conceitos genéticos relevantes, como polimorfismos de nucleotídeo único e expressão gênica, além dos pilares de banco de dados. As funcionalidades desenvolvidas foram testadas com dados reais, demonstrando eficiência na inserção em massa via arquivos CSV e na exportação para formatos compatíveis com ferramentas como R, Python e softwares de análise genômica. A estrutura de permissões por grupos foi validada, permitindo controle de acesso conforme o perfil do usuário.

Palavras-chave: Banco de dados. Genética. Django.

ABSTRACT

Chromosomal and genomic analysis has become a fundamental tool for personalized diagnosis and treatment in clinical medicine. This work proposes the development of a platform for storage, organization, and analysis of genetic and phenotypic data, meeting the demand of the Graduate Program in Genetics at UFPR in partnership with LAInfS. The main motivation arises from the need to structure data from previous research stored in spreadsheets, ensuring greater integrity, standardization, and security. To this end, a relational database was modeled based on spreadsheets provided by professors and researchers, incorporating functionalities such as importing and exporting data in CSV format, user management, and access control. The platform was developed in Python, using the Django framework. The work also presents a review of relevant genetic concepts, such as single nucleotide polymorphisms and gene expression, as well as the pillars of databases. The developed functionalities were tested with real data, demonstrating efficiency in bulk insertion via CSV files and in exporting to formats compatible with tools like R, Python, and genomic analysis software. The group permission structure was validated, allowing access control according to the user's profile.

Keywords: Database. Genetics. Django.

LISTA DE FIGURAS

1	FIGURA 1 – DIAGRAMA DE CASO DE USO	16
2	FIGURA 2 – DIAGRAMA DE ESQUEMA BANCO DE DADOS CANTA SABIÁ	17
3	FIGURA 3 – DIAGRAMA DE ESQUEMA BANCO DE DADOS ALPE	18
4	FIGURA 4 – DIAGRAMA DE ESQUEMA BANCO DE DADOS PROJETO RATOS	18
5	FIGURA 5 – DIAGRAMA ER CANTA SABIÁ.	20
6	FIGURA 6 – DIAGRAMA ER BANCO ALPE	21
7	FIGURA 7 – DIAGRAMA ER PROJETO RATOS	22
8	FIGURA 8 – FLUXO DE CONTROLE MODEL VIEW TEMPLATE	26
9	FIGURA 9 – PÁGINA DE LOGIN.	31
10	FIGURA 10 – PÁGINA INICIAL	32
11	FIGURA 11 – PÁGINA GERENCIAMENTO DAS INFORMAÇÕES	32
12	FIGURA 12 – TELA IMPORTAÇÃO - ESCOLHE BANCO	33
13	FIGURA 13 – TELA IMPORTAÇÃO - ESCOLHE TABELA	34
14	FIGURA 14 – TELA IMPORTAÇÃO - ESCOLHE ARQUIVO.	34
15	FIGURA 15 – TELA EXPORTAÇÃO - ESCOLHE BANCO	35
16	FIGURA 16 – TELA EXPORTAÇÃO - ESCOLHE TABELA	35
17	FIGURA 17 – TELA GERENCIAMENTO DE USUÁRIO	36
18	FIGURA 18 – TELA CRIAÇÃO DE USUÁRIO.	36
19	FIGURA 19 – TELA PARA CRIAÇÃO E CONFIGURAÇÃO DE UM GRUPO .	37

LISTA DE TABELAS

1	TABELA 1 – REQUISITOS FUNCIONAIS	15
2	TABELA 2 – REQUISITOS NÃO-FUNCIONAIS.	15
3	TABELA 3 – DESCRIÇÃO DAS ENTIDADES DO BANCO DE DADOS CANTA SABIÁ.	24
4	TABELA 4 – DESCRIÇÃO DOS OBJETIVOS DE CADA ENTIDADE	26

LISTA DE ACRÔNIMOS

API	Application Programming Interface
CSS	Cascading Style Sheets
CSV	Valores Separados por Vírgula
DNA	Ácido Desoxirribonucleico
DTL	Django Template Language
GWAS	Estudo de Associação Genômica Ampla
LAInfS	Liga Acadêmica de Informática em Saúde
MVT	Model-View-Template
NLM	National Library of Medicine
RNA	Ácido Ribonucleico
RNA-m	RNA mensageiro
SNP	Polimorfismo de Nucleotídeo Único
SMTP	Simple Mail Transfer Protocol
UFPR	Universidade Federal do Paraná
UTF-8	Unicode Transformation Format 8-bit

SUMÁRIO

1	INTRODUÇÃO	10
1.1	JUSTIFICATIVA	10
1.2	OBJETIVOS	11
1.2.1	Objetivos Gerais	11
1.2.2	Objetivos Específicos	11
2	REVISÃO DE LITERATURA	12
2.1	BANCO DE DADOS	12
2.1.1	Banco de dados relacional	12
2.1.2	Modelo entidade-relacional	12
2.2	PRINCIPAIS CONCEITOS GENÉTICOS	13
2.2.1	Estrutura do DNA	13
2.2.2	Polimorfismo de nucleotídeo único	13
2.2.3	Expressão Gênica	13
2.3	IMPORTÂNCIA DOS SNPS PARA PESQUISAS	14
3	MATERIAIS E MÉTODOS	15
3.1	REQUISITOS DO SISTEMA	15
3.1.1	Requisitos funcionais	15
3.1.2	Requisitos não-funcionais	15
3.2	DIAGRAMA DE CASO DE USO	16
3.3	DIAGRAMA DE ESQUEMA	17
3.4	DIAGRAMA ENTIDADE-RELACIONAMENTO	19
3.5	NORMALIZAÇÃO	22
3.5.1	Primeira Forma Normal	22
3.5.2	Segunda Forma Normal	22
3.5.3	Terceira Forma Normal	23
3.6	IMPLEMENTAÇÃO DA PLATAFORMA	23
3.7	MODELAGEM DE DADOS	23
3.7.1	Modelagem banco Canta Sabiá	23
3.7.2	Modelagem banco Alpe	24
3.7.3	Modelagem banco Projeto Ratos	25
3.8	DJANGO	26
3.8.1	Model	26
3.8.2	View	27
3.8.3	Template	27

3.9	INTERFACE COM HTML, CSS E INTEGRAÇÃO COM DJANGO	27
3.10	CRIAÇÃO DOS BANCOS DE DADOS	28
3.11	IMPLEMENTAÇÃO DAS PRINCIPAIS FUNCIONALIDADES	28
3.11.1	Implementação das funcionalidades gerenciamento de usuários, controle de grupos e alteração de credenciais de acesso.	28
3.11.2	Implementação do acesso aos dados	29
3.11.3	Implementação da importação e exportação dos dados.	29
4	RESULTADOS E DISCUSSÕES.	31
5	CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS.	38
	REFERÊNCIAS	39

1 INTRODUÇÃO

A análise cromossômica e genômica tem se consolidado como uma ferramenta diagnóstica crucial na medicina clínica, com aplicações abrangendo desde o diagnóstico de condições médicas associadas a alterações cromossômicas e genéticas até o desenvolvimento de tratamentos personalizados. De acordo com Thompson e Thompson (2016), essa área envolve, entre outros aspectos, a identificação de genes específicos e a compreensão de seus papéis em doenças, a genômica do câncer e a avaliação da integridade genética para tratamentos terapêuticos.

Este trabalho propõe o desenvolvimento de uma plataforma destinada ao armazenamento e acesso aos dados provenientes dos projetos realizados pelo Departamento de Genética da UFPR. A iniciativa nasce de uma colaboração entre o Programa de Pós-Graduação em Genética da UFPR e a Liga de Pesquisadores Acadêmica de Informática em Saúde (LAIInfS), buscando suprir uma demanda real pela estruturação, armazenamento e gerenciamento de bancos de dados para projetos. Os bancos de dados foram elaborados com base em pesquisas realizadas, como os estudos sobre genes relacionados à linguagem e musicalidade no projeto Canta Sabiá, e as mutações no gene BCHE no projeto Alpe, além de estruturarem o modelo para pesquisas ainda em andamento, como o projeto Ratos, que visa fazer a análise da expressão gênica e da atividade enzimática em ratos.

1.1 JUSTIFICATIVA

A motivação para este projeto surgiu da necessidade de criar estruturas de dados para projetos em andamento, bem como de organizar os dados de projetos anteriores, armazenados em planilhas do Excel pelo Departamento de Genética. Embora planilhas sejam eficientes para análises pontuais, sistemas de banco de dados estruturados oferecem vantagens significativas, como maior integridade dos dados, organização e controle de acesso.

A organização dos dados é essencial para uma análise eficaz e para alcançar resultados mais precisos. As pesquisas anteriores estavam armazenadas em planilhas com relacionamentos não definidos, além de apresentarem atributos com dados não padronizados, ou seja, diferentes tipos de informações para o mesmo atributo, prejudicando a compreensão e a análise dos dados.

Quando os dados estão dispersos em diversos arquivos e formatos, aumenta a complexidade no desenvolvimento de novos programas que consigam recuperar e utilizar essas informações de forma eficiente. Esse desafio, conhecido como isolamento dos dados, dificulta o reaproveitamento e a integração das informações armazenadas (Silberschatz et al., 2020).

Outro ponto crítico é a garantia da integridade dos dados. É necessário que os valores armazenados satisfaçam determinadas restrições de consistência. Conforme observado por Silberschatz et al. (2020), quando essas regras são inseridas diretamente nos códigos das aplicações, torna-se difícil modificá-las ou expandi-las, especialmente quando envolvem dados distribuídos em diferentes arquivos, comprometendo a confiabilidade do sistema.

Para os projetos em andamento e que não possuíam nenhuma estrutura implementada, foi iniciado o processo de coleta já com tabelas organizadas. Essa pré-implementação irá garantir que seja possível adicionar informações complementares conforme necessário, mantendo uma estrutura de banco de dados definida, o que facilitará a alimentação e a expansão futura.

Além disso, a análise desses dados é, na maioria das vezes, realizada por meio da linguagem R. Com a organização adequada da origem dos dados, esse processo será

significativamente aprimorado, pois reduzirá o tempo gasto com a limpeza de dados e facilitará a compreensão das informações e dos relacionamentos das tabelas.

1.2 OBJETIVOS

1.2.1 Objetivos Gerais

Armazenamento, organização de dados genéticos e fenotípicos associados a doenças específicas ou tratamentos farmacogenéticos. A ferramenta visa facilitar o acesso seguro e eficiente a essas informações, otimizando pesquisas científicas no Setor de Ciências Biológicas da UFPR e contribuindo para o uso de dados para os projetos futuros.

1.2.2 Objetivos Específicos

- Modelar e implementar um banco de dados relacional estruturado com base em planilhas de pesquisa fornecidas por docentes e pesquisadores da UFPR, além de dados provenientes de projetos em andamento, assegurando a padronização, a integridade e a interoperabilidade dos dados genéticos e fenotípicos.
- Desenvolver funcionalidades de registro e atualização de dados, como polimorfismos de nucleotídeo único (SNPs, *Single Nucleotide Polymorphism*), atividade enzimática, expressão gênica, genes associados e informações diagnósticas, permitindo a inserção de novos registros ou deleção.
- Facilitar a importação de novos dados por meio de planilhas no formato CSV.
- Permitir a exportação dos dados armazenados no banco de dados.
- Fornecer ao administrador a capacidade de gerenciar o acesso ao banco de dados por meio de controle de usuários.

O presente trabalho está estruturado da seguinte forma: o capítulo de Revisão de Literatura aborda os conceitos teóricos que fundamentam o projeto, englobando os princípios de banco de dados e de genética. Em seguida, o capítulo de Materiais e Métodos descreve detalhadamente o processo de desenvolvimento da plataforma, desde a especificação dos requisitos e a modelagem dos dados até as tecnologias utilizadas. O capítulo de Resultados e Discussões apresenta a plataforma finalizada. Por fim, as Considerações Finais apresentam as contribuições do trabalho e propõem melhorias futuras para a plataforma.

2 REVISÃO DE LITERATURA

2.1 BANCO DE DADOS

Com o crescimento das informações armazenadas, os sistemas tradicionais baseados em arquivos passaram a apresentar limitações significativas. Entre os principais problemas estavam a separação e o isolamento dos dados, o que dificultava a realização de consultas complexas, além da redundância, inconsistência e problemas de segurança. A duplicação frequente de dados e a incompatibilidade entre formatos de arquivos também contribuem para a ineficiência dos sistemas, especialmente diante do aumento exponencial da informação (Cardoso e Cardoso, 2012).

Um banco de dados é uma coleção organizada de dados que possuem coerência e significado, sendo criado com um propósito específico e voltado para atender às necessidades de determinados usuários e aplicações. Ele representa uma parte do mundo real, por isso, qualquer mudança nesse contexto é refletida automaticamente no banco de dados (Machado, 2024). No contexto do armazenamento de informações, os bancos de dados são compreendidos não apenas como simples repositórios, mas como estruturas organizadas a partir de modelos, que incorporam recursos para facilitar tanto o armazenamento quanto a consulta dos dados (Cardoso e Cardoso, 2012).

2.1.1 Banco de dados relacional

Um banco de dados relacional é composto por uma coleção de tabelas, cada uma com um nome exclusivo. Cada tabela é organizada em colunas, que representam os atributos dos dados, e linhas, que armazenam os registros individuais. As colunas possuem nomes específicos e cada linha corresponde a um campo de dados, representando um conjunto de valores relacionados entre si. Esse modelo permite organizar e acessar as informações de forma estruturada, utilizando chaves para identificar unicamente cada registro (Silberschatz et al., 2020).

Entre essas chaves, destaca-se a chave primária, que é utilizada para identificar de forma única os registros em uma tabela. Além de garantir unicidade, ela funciona como um índice, organizando automaticamente os dados ao definir sua ordem física. A chave primária pode ser composta por um ou mais campos, dependendo da complexidade da estrutura da tabela (William Pereira Alves, 2011).

Complementando esse mecanismo, a chave estrangeira é o campo responsável por estabelecer ligações entre tabelas distintas. Ela se refere à chave primária de outra tabela, permitindo a criação de relações lógicas entre conjuntos de dados. Mantém seu papel identificador, atuando como referência externa. (Cardoso e Cardoso, 2012).

2.1.2 Modelo entidade-relacional

O modelo entidade-relacional (E-R) representa a estrutura lógica geral de um banco de dados. Foi desenvolvido para facilitar a execução de projetos de bancos de dados, sendo capaz de representar de maneira gráfica a estrutura das informações. O modelo E-R baseia-se em três noções fundamentais: conjuntos de entidades, conjuntos de relacionamentos e atributos. (Silberschatz et al., 2020).

Entidade pode ser definida por um objeto do mundo real que é diferenciável de todos os outros. Possui propriedades e características que a identificam. Os conjuntos de entidades

englobam entidades que compartilham propriedades e características em comum (Silberschatz et al., 2020). As propriedades e características de uma entidade são definidas pelos atributos.

O relacionamento em um modelo Entidade-Relacionamento representa um conjunto de associações entre instâncias de diferentes entidades. Ele descreve como os dados estão conectados, ou seja, como uma entidade se relaciona com outra dentro do domínio do sistema. Esses dois componentes são interdependentes e refletem ações que ocorrem entre os elementos do sistema, o que impossibilita a sua análise de forma isolada (ALVES, 2014).

2.2 PRINCIPAIS CONCEITOS GENÉTICOS

2.2.1 Estrutura do DNA

A estrutura do DNA é fundamental para a compreensão do genoma humano, e sua análise começa pela descrição de suas unidades básicas. O DNA é uma macromolécula polimérica composta por três componentes principais: a desoxirribose (um açúcar de cinco carbonos), uma base nitrogenada e um grupo fosfato. As bases nitrogenadas podem ser purinas, como adenina (A) e guanina (G), ou pirimidinas, como timina (T) e citosina (C). Esses nucleotídeos, que consistem em uma base, um fosfato e uma fração de açúcar, se ligam por meio de ligações fosfodiéster, formando longas cadeias de DNA. No caso do genoma humano, essas cadeias se organizam em uma dupla hélice, composta por duas fitas complementares, cujas bases se emparelham de maneira específica (A com T e G com C). Essa organização permite a replicação exata da informação genética e sua transmissão das células originais para as células-filhas, garantindo a estabilidade genética ao longo das gerações (Thompson e Thompson, 2016).

2.2.2 Polimorfismo de nucleotídeo único

Polimorfismos de nucleotídeo único (SNPs) são o tipo mais comum de variação genética entre os seres humanos. Cada SNP representa uma diferença em um único bloco de construção do DNA, chamado nucleotídeo (MedlinePlus, 2025). A classificação de uma variante genética como polimorfismo não depende de sua natureza molecular, do seu impacto funcional ou de sua localização no gene, mas exclusivamente de sua frequência na população, sendo necessário que ocorra em mais de 1% dos alelos (Thompson e Thompson, 2016).

Aproximadamente uma em cada mil bases do genoma humano apresenta variação, sendo os SNPs os mais comuns (Turnpenney e Ellard, 2021). Esses polimorfismos, geralmente bialélicos, podem ocorrer tanto em regiões codificadoras quanto não codificadoras do DNA. Embora muitas dessas variações estejam situadas em regiões não codificantes, algumas podem afetar diretamente proteínas e contribuir para a diversidade genética entre indivíduos e populações humanas (Thompson e Thompson, 2016).

2.2.3 Expressão Gênica

A expressão gênica refere-se ao processo pelo qual informações contidas no DNA são utilizadas para produzir RNA e polipeptídeos, de forma distinta de uma célula para outra. Embora o DNA tenha a função geral de armazenar e transmitir informações genéticas, além de apoiar funções cromossômicas, ele também possui funções específicas que variam conforme o tipo celular. Genes, que são segmentos de DNA distribuídos de maneira irregular ao longo de sua sequência, servem como moldes para a produção de RNA por meio da transcrição. O RNA transcrito passa por etapas de maturação que podem resultar em diferentes tipos de RNA, como RNAs mensageiros ou não codificantes, que desempenham funções essenciais nas células.

Alguns produtos gênicos são necessários para todas as células, enquanto outros são específicos de tipos celulares ou até de uma célula individual, como ocorre em linfócitos B e T. A diferença entre as células não está na composição de seu DNA, mas sim na expressão gênica, que varia de acordo com as necessidades celulares, com alguns genes sendo expressos em todas as células e outros apresentando expressão específica em determinados tecidos ou fases do ciclo celular (Strachan e Read, 2013).

2.3 IMPORTÂNCIA DOS SNPS PARA PESQUISAS

Os SNPs são as variações genéticas mais comuns no genoma humano e desempenham um papel essencial no mapeamento de suscetibilidades às doenças, identificação de predisposições genéticas e resposta a medicamentos (Thompson e Thompson, 2016). Sua análise é fundamental para desvendar a base genética de condições complexas e impulsionar a medicina de precisão. Aproximadamente 90% da variação genética humana está relacionada aos SNPs, que ocorrem com frequência acima de 1% (Sunyaev et al., 2000). A identificação e a caracterização de SNPs comuns e biologicamente relevantes, especialmente aqueles associados ao risco de enfermidades, são etapas essenciais para que esses marcadores possam ser utilizados de forma ampla como ferramentas genéticas. Considerando que as frequências alélicas dos SNPs variam entre diferentes populações e grupos étnicos, iniciativas internacionais optaram por utilizar painéis geneticamente diversos, ampliando, assim, as chances de descoberta de variantes informativas (The International SNP Map Working Group, 2001). Nos últimos cinco anos, os SNPs foram os principais marcadores genéticos de mais de 21 mil publicações feitas no PubMed, mantido pela *National Library of Medicine (NLM)*. A partir disso, é possível notar a consolidação dos SNPs como instrumentos de identificação de locais de risco, na construção de *scores* poligênicos preditivos e no avanço da medicina personalizada. O uso dos SNPs como principais marcadores genéticos também se estende à farmacogenômica, biologia evolutiva e epidemiologia genética.

3 MATERIAIS E MÉTODOS

3.1 REQUISITOS DO SISTEMA

A engenharia de requisitos é uma etapa crítica no desenvolvimento de sistemas de informação robustos. Ela deve abranger:

- **Requisitos funcionais:** definem as funcionalidades essenciais que o sistema deve executar, como cadastro, consulta e processamento de dados (Pressman e Maxim, 2020).
- **Requisitos não-funcionais:** referem-se a atributos que incluem aspectos como qualidade, desempenho, segurança e restrições gerais do sistema (Pressman e Maxim, 2020).

As Tabelas 1 e 2 resumem, respectivamente, os requisitos funcionais e não funcionais da plataforma.

3.1.1 Requisitos funcionais

TABELA 1 – REQUISITOS FUNCIONAIS

Código	Requisito
RF-01	Gerenciamento do banco de dados, incluindo adição, remoção e edição dos dados.
RF-02	Importação de dados, com validação de estrutura e tipagem dos campos.
RF-03	Cadastro de novos usuários pelo admin com campos obrigatórios: nome, e-mail institucional e nível de acesso.
RF-04	Hierarquia de privilégios (admin, pesquisador) com permissões diferenciadas para cada nível e com acessos de bancos diferentes.
RF-05	Permitir que usuários administradores criem grupos personalizados e definam perfis de acesso padrão para cada grupo.
RF-06	Permitir que o usuário recupere a senha pelo seu e-mail de cadastro.
RF-07	Permitir a exportação do banco de dados.

FONTE: O Autor (2025)

3.1.2 Requisitos não-funcionais

TABELA 2 – REQUISITOS NÃO-FUNCIONAIS

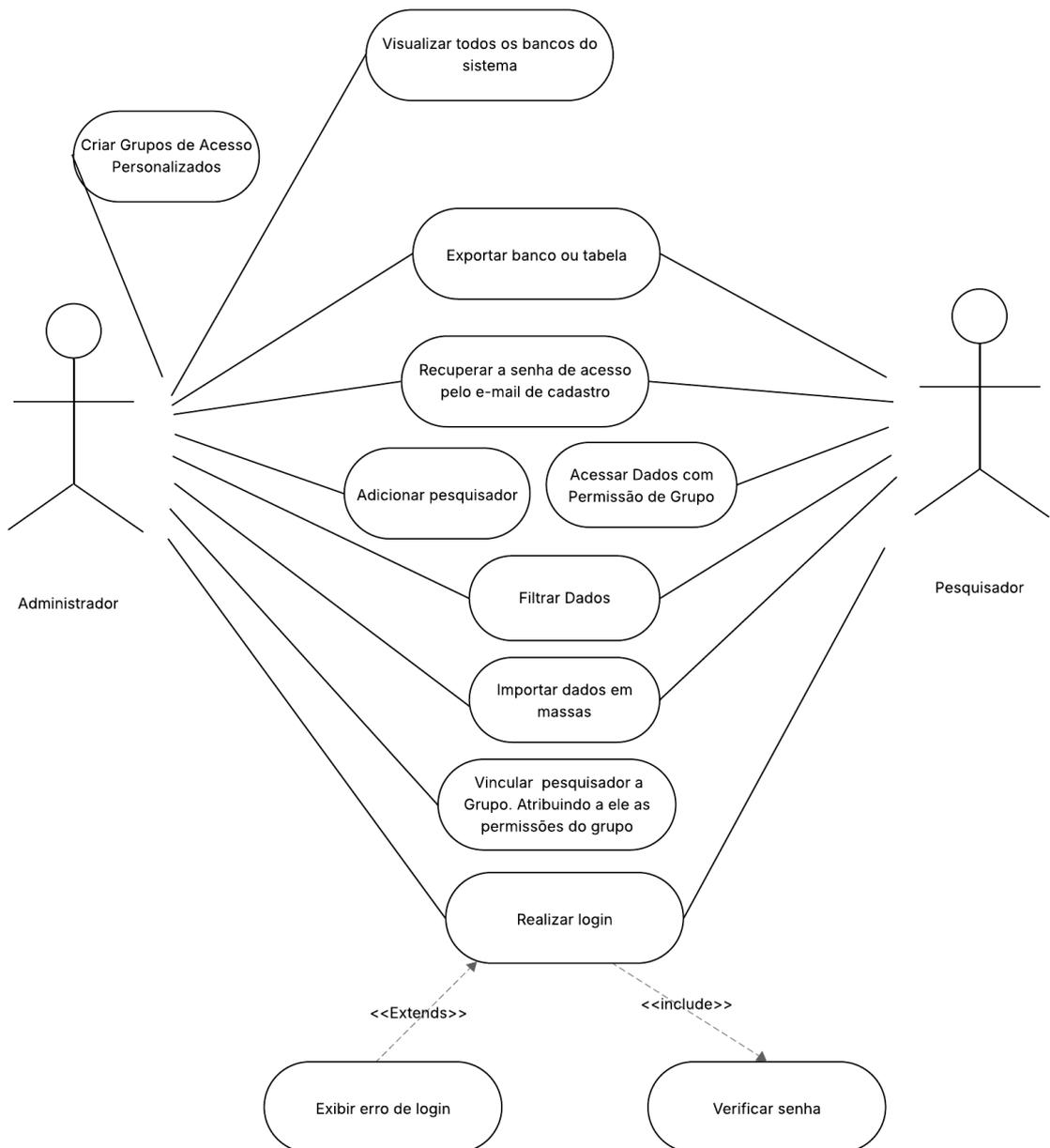
Código	Requisito
RNF-01	A plataforma deve ser simples e intuitiva, facilitando o aprendizado inicial e promovendo a memorização eficiente de suas funcionalidades pelos usuários.
RNF-02	A plataforma deve ser capaz de se conectar com outros Sistemas de Gerenciamento de Banco de Dados, sendo ele PostgreSQL, MariaDB, MySQL ou Oracle.
RNF-03	A plataforma deve permitir importação e exportação dos dados no formato CSV.
RNF-04	A plataforma deve ser desenvolvida em Python, utilizando o framework Django.
RNF-05	A plataforma deve exigir autenticação de usuário com nome de usuário e senha.

FONTE: O Autor (2025)

3.2 DIAGRAMA DE CASO DE USO

Segundo Sommerville (2011), os casos de uso são representados por diagramas que descrevem, de forma visual, todas as possíveis interações entre o sistema e seus atores, os quais podem ser usuários humanos ou outros sistemas. Essas interações são organizadas em elipses, conectadas aos atores por linhas, podendo incluir setas para indicar o fluxo de interação. A seguir, apresenta-se o diagrama de caso de uso da plataforma, que ilustra suas principais interações com os usuários.

FIGURA 1 – DIAGRAMA DE CASO DE USO



FONTE: O autor (2025)

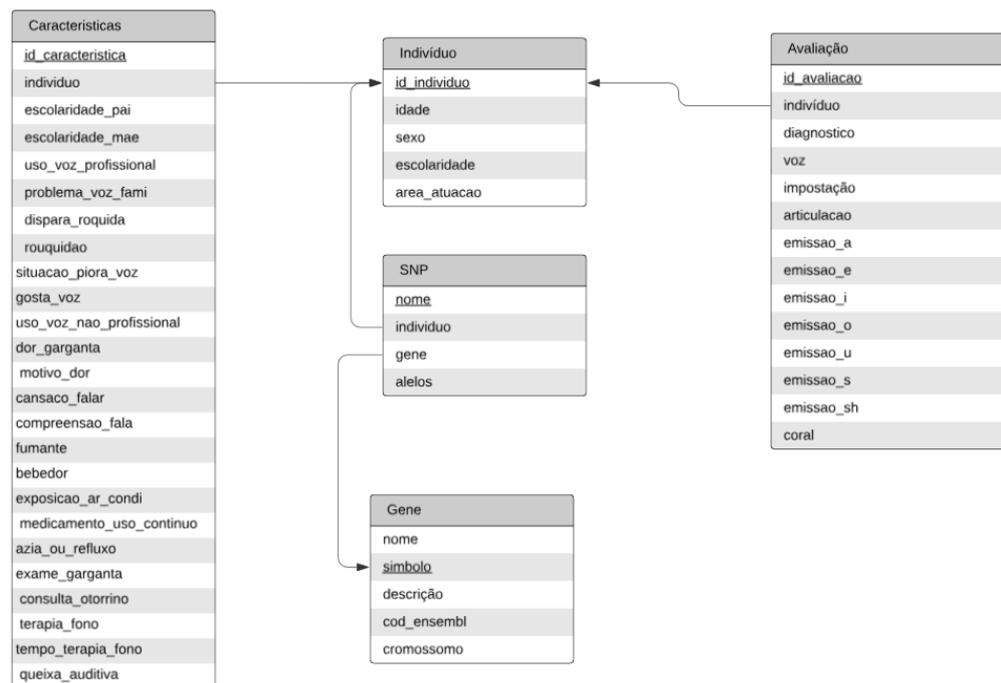
A Figura 1 apresenta os dois atores principais do sistema: o Administrador e o Pesquisador. Ambos possuem as funcionalidades de exportação de bancos de dados, filtro

de dados, importação de dados em massa, recuperação de senhas e realização de *login*. O Administrador possui um conjunto de funcionalidades com um nível maior de controle quando comparado ao Pesquisador. Dentre elas estão a criação de grupos de acesso personalizados, a visualização de todos os bancos do sistema, adição de novos pesquisadores e a vinculação de pesquisadores a grupos, atribuindo-lhes permissões específicas. Por outro lado, o Pesquisador possui um nível de controle menor e consegue interagir com o sistema para acessar dados com permissão de grupo. Essa estrutura demonstra a hierarquia de acesso e as capacidades de cada perfil de usuário dentro da plataforma.

3.3 DIAGRAMA DE ESQUEMA

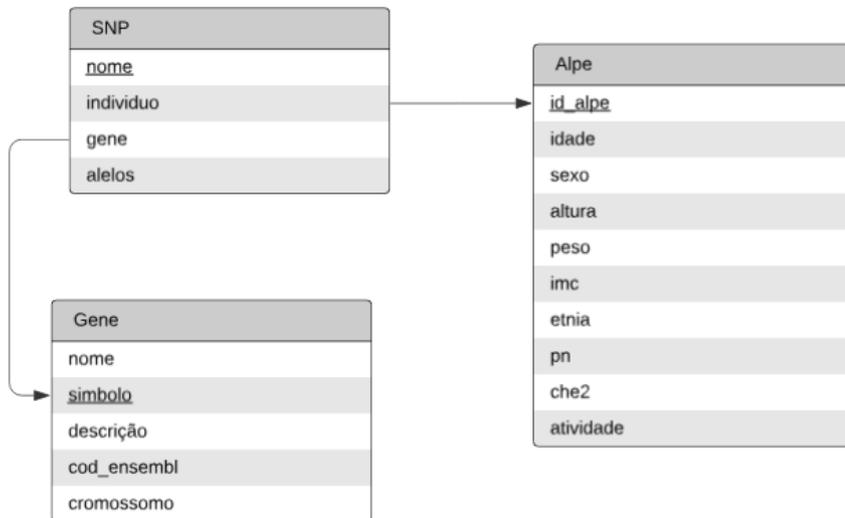
Os diagramas a seguir representam o esquema do banco de dados, evidenciando as tabelas e seus relacionamentos por meio de chaves primárias e chaves estrangeiras. Para facilitar a compreensão visual, optou-se por uma versão simplificada, que destaca a estrutura principal sem exibir todos os atributos.

FIGURA 2 – DIAGRAMA DE ESQUEMA BANCO DE DADOS CANTA SABIÁ



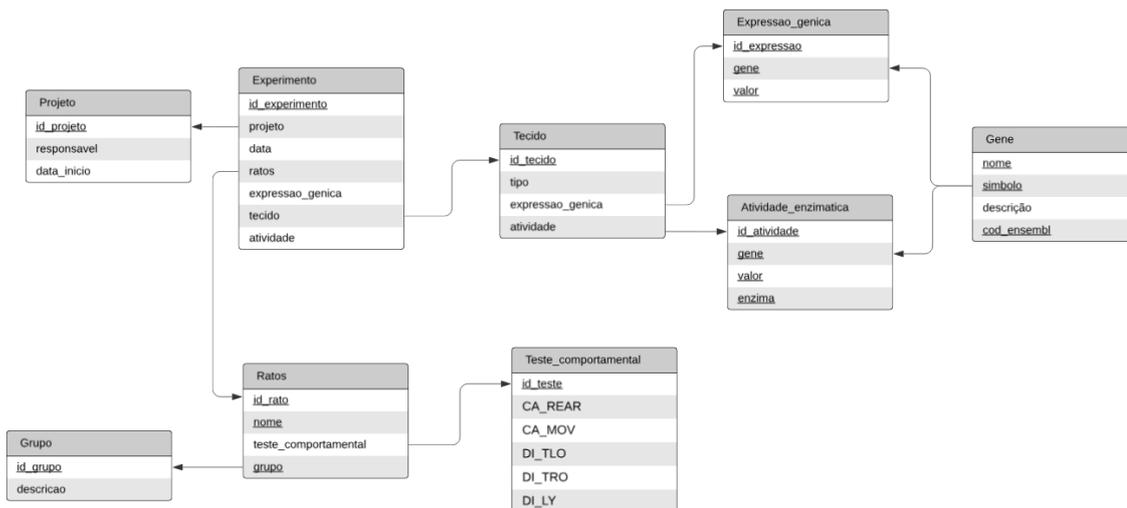
FONTE: O autor (2025)

FIGURA 3 – DIAGRAMA DE ESQUEMA BANCO DE DADOS ALPE



FONTE: O autor (2025)

FIGURA 4 – DIAGRAMA DE ESQUEMA BANCO DE DADOS PROJETO RATOS



FONTE: O Autor (2025)

3.4 DIAGRAMA ENTIDADE-RELACIONAMENTO

Os diagramas Entidade-Relacionamento (E/R) constituem uma técnica visual que facilita a compreensão e a comunicação dos principais elementos de um banco de dados, sendo essa representação gráfica um dos principais fatores responsáveis pela popularidade do modelo E/R (Elmasri e Navathe, 2005). Uma entidade representa algo do mundo real ou do universo em estudo, caracterizando-se por atributos, enquanto os relacionamentos descrevem interações entre entidades (ALVES, 2014).

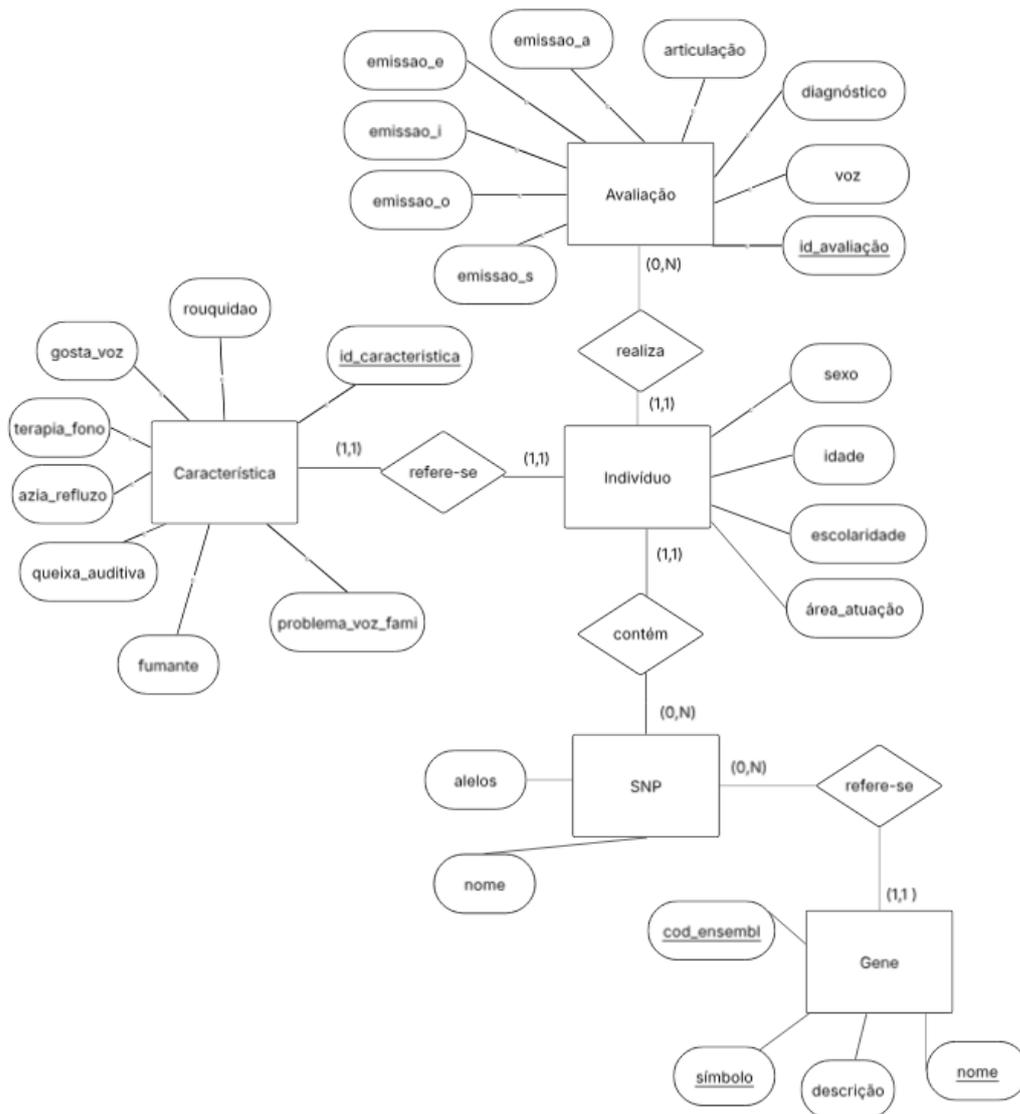
No diagrama também são representadas as cardinalidades dos relacionamentos. As cardinalidades de mapeamento indicam quantas instâncias de uma entidade podem se relacionar com instâncias de outra, sendo especialmente úteis para representar relacionamentos entre múltiplos conjuntos de entidades (Silberschatz et al., 2020).

Existem as seguintes cardinalidades:

- **Um para um:** Cada item A se liga a, no máximo, um item de B, e o contrário também.
- **Um para muitos:** Um item A pode se ligar a vários de B, mas cada item de B se liga a, no máximo, um de A.
- **Muitos para um:** Vários itens A podem se ligar a um único de B, mas cada item de A só se liga a um de B.
- **Muitos para muitos:** Itens A e B podem se ligar a vários itens entre si.

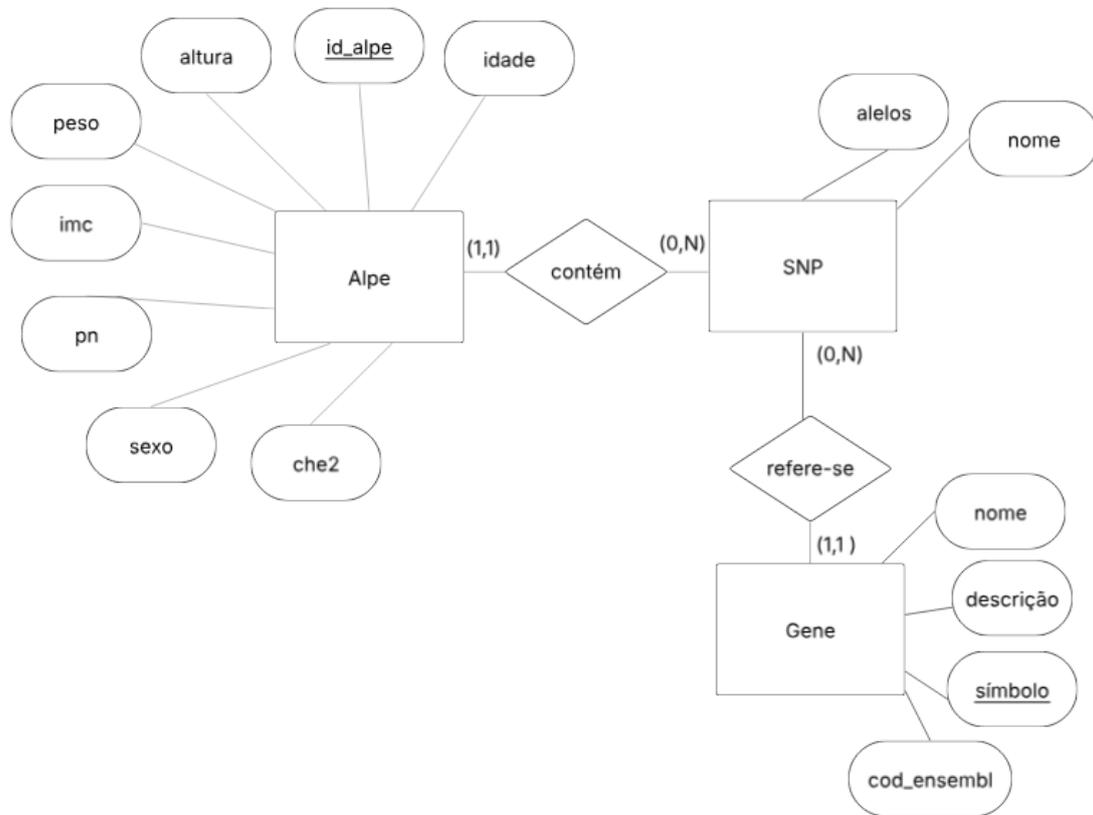
A seguir, a Figura 5, 6 e 7 representa um Diagrama Entidade-Relacionamento (DER) correspondente ao seu respectivo banco de dados. Nestes diagramas, as entidades são representadas por triângulos, os relacionamentos por retângulos e os atributos por elipses. Os atributos que representam chaves primárias estão sublinhados, conforme a notação utilizada. Além disso, as cardinalidades estão indicadas junto aos relacionamentos, especificando o número mínimo e máximo de ocorrências entre as entidades associadas.

FIGURA 5 – DIAGRAMA ER CANTA SABIÁ



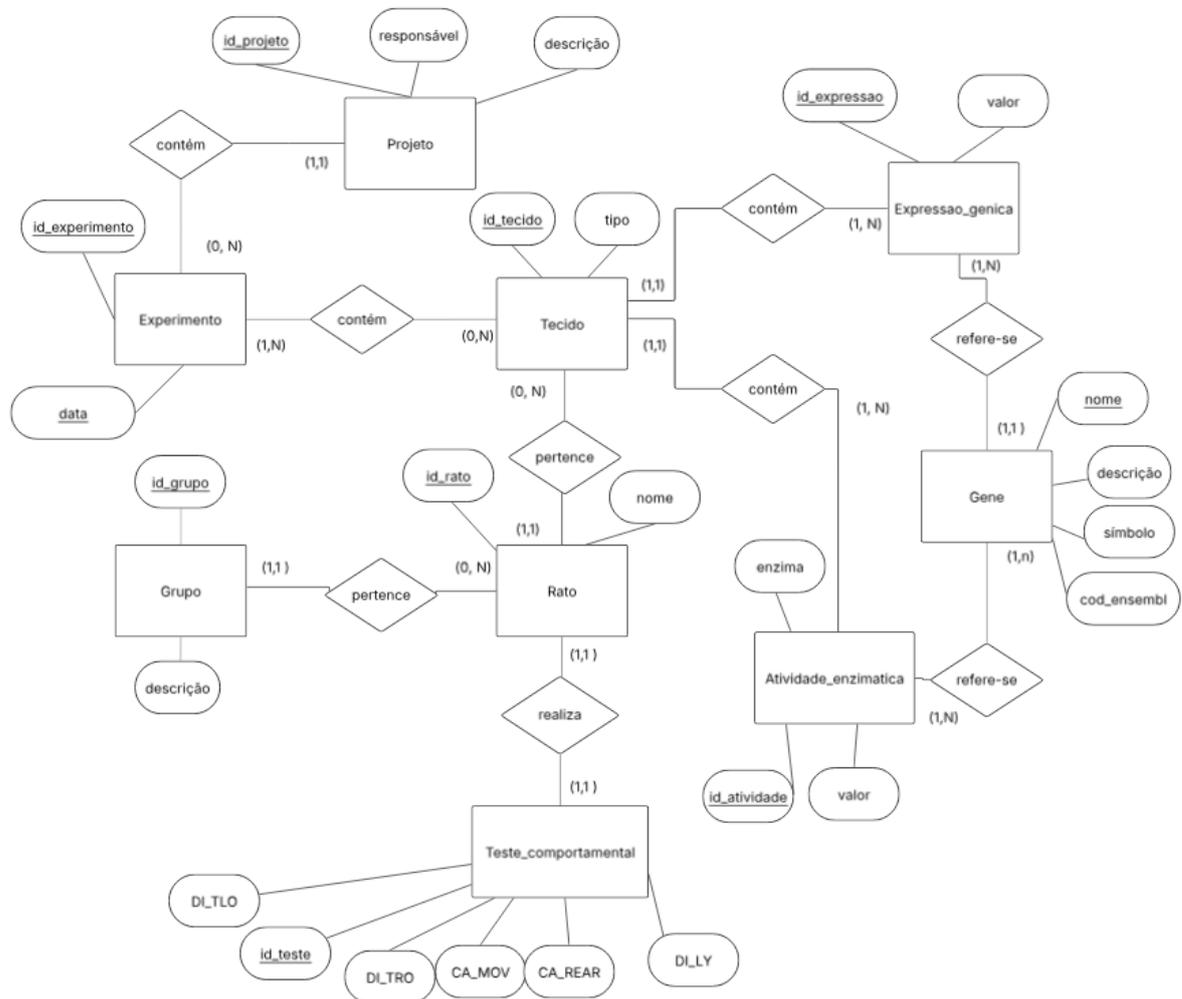
FONTE: O autor (2025)

FIGURA 6 – DIAGRAMA ER BANCO ALPE



FONTE: O autor (2025)

FIGURA 7 – DIAGRAMA ER PROJETO RATOS



FONTE: O Autor (2025)

3.5 NORMALIZAÇÃO

De acordo com Silberschatz et al. (2020), a normalização é um método essencial para o projeto de bancos de dados relacionais. Este método visa eliminar redundâncias e facilitar a recuperação de informações. Para isso, os autores destacam a importância de estruturar os dados em formas normais adequadas, utilizando dependências funcionais como base para a análise.

3.5.1 Primeira Forma Normal

A Primeira Forma Normal (1FN) estabelece que os atributos devem conter valores atômicos, ou seja, indivisíveis e únicos. Além disso, não devem existir atributos multivalorados ou compostos, e os valores que compõem a chave primária não podem se repetir (ALVES, 2014).

3.5.2 Segunda Forma Normal

Uma tabela está na Segunda Forma Normal (2FN) quando todos os seus atributos não-chave dependem funcionalmente de toda a chave primária, especialmente no caso de chaves compostas. Isso significa que não deve haver dependência de apenas parte da chave, mas sim uma

dependência funcional total, garantindo a integridade dos dados e evitando redundâncias parciais (Cardoso e Cardoso, 2012). No projeto desenvolvido, não há utilização de chaves primárias compostas, o que facilitou o cumprimento dessa norma.

3.5.3 Terceira Forma Normal

A Terceira Forma Normal (3FN) estabelece que todos os atributos não pertencentes à chave primária devem depender exclusivamente dela, de forma direta e funcional. Dessa forma, elimina-se qualquer tipo de dependência transitiva entre atributos, o que é feito por meio da reorganização da tabela, garantindo que cada dado esteja vinculado diretamente à chave primária, sem redundâncias estruturais (Cardoso e Cardoso, 2012).

3.6 IMPLEMENTAÇÃO DA PLATAFORMA

Inicialmente, conduziu-se uma fase de planejamento envolvendo reuniões com o coordenador do trabalho para análise dos dados disponíveis e definição dos requisitos. Nesta etapa, elaborou-se em conjunto o modelo relacional de um dos bancos de dados, estabelecendo as entidades, seus atributos e os relacionamentos.

Posteriormente, deu-se início à implementação técnica. Para o *back-end*, optou-se pela linguagem Python em conjunto com o *framework Django*, visando à criação de modelos semelhantes ao modelo relacional previamente definido, a implementação da lógica de negócios nas *views* e a configuração da interação com a interface. O *front-end* foi desenvolvido utilizando HTML, CSS e o *framework Bootstrap*.

É importante destacar que o desenvolvimento deste projeto não seguiu uma ordem estritamente sequencial. Inicialmente, foi estruturado apenas um dos bancos de dados. Somente após o desenvolvimento da plataforma de acesso, os outros dois bancos foram implementados. A seguir, serão detalhados o processo de modelagem e criação dos bancos de dados e o desenvolvimento da plataforma.

3.7 MODELAGEM DE DADOS

A estruturação dos bancos de dados foi baseada em pesquisas realizadas e armazenadas em tabelas fornecidas pelo Laboratório de Polimorfismos e Ligação do Departamento de Genética e por pesquisas que ainda estão em andamento e que não possuíam os dados estruturados. Para auxiliar na modelagem e visualização dos relacionamentos entre as tabelas, foi utilizada a ferramenta Lucidchart para desenvolver o Diagrama de Classe e o Modelo entidade-relacional. Reuniões realizadas com o vice-coordenador do programa de pós-graduação, foram essenciais para o entendimento dos dados e para a validação das estruturas desenvolvidas.

3.7.1 Modelagem banco Canta Sabiá

O banco Canta Sabiá é o resultado de alguns projetos que visam relacionar os genes humanos, a linguagem e a musicalidade. Essa relação se dá pelos genes, DRD2, FOXP2, DRD4, AVPR1A, ITGB3, SLC6A4, COMT. De acordo com Antonio et al. (2024), polimorfismos em genes do sistema dopaminérgico, como COMT (VAL158MET), DRD2 (rs6278) e DRD4 (VNTR), estão associados a características cognitivas, incluindo inteligência, memória, dificuldades emocionais e dependência, além de apresentarem relação com aspectos musicais, como reconhecimento de altura e improvisação. Além deles, o gene FOXP2 codifica um fator de transcrição da família *forkhead box* (FOX), expresso no cérebro fetal e adulto, além de outros órgãos, como pulmões

e intestino. Esse gene é essencial para o desenvolvimento das regiões cerebrais relacionadas à fala e à linguagem durante a embriogênese, e mutações nele podem causar transtornos de fala e linguagem (NCBI RefSeq, 2010).

Visando relacionar aspectos fenotípicos e ambientais dos indivíduos aos genes associados à linguagem e musicalidade, foi realizada uma pesquisa que considerou dados pessoais como sexo, idade, escolaridade e área de atuação, dados de características relacionadas à prática musical e à fala, complementadas por avaliações auditivas dos participantes. Para apoiar essa análise, foram armazenados os SNPs relacionados ao tema, possibilitando uma investigação das interações entre fatores genéticos, ambientais e fenotípicos no desenvolvimento das habilidades linguísticas e musicais. A Tabela 3 apresenta a descrição de cada entidade do banco de dados Canta Sabiá. Ademais, os diagramas exibidos nas seções 3.3 e 3.4 também detalham a estrutura final do modelo.

TABELA 3 – DESCRIÇÃO DAS ENTIDADES DO BANCO DE DADOS CANTA SABIÁ

Tabela	Descrição
Indivíduo	Armazena informações pessoais dos participantes da pesquisa, como sexo, idade, escolaridade e área de atuação.
SNPs	Contém os polimorfismos de nucleotídeo único relacionados aos genes estudados, fundamentais para análises genéticas associadas à linguagem e musicalidade.
Gene	Registra dados sobre os genes relacionados à pesquisa.
Avaliação	Armazena os resultados das avaliações auditivas e outros testes realizados para mensurar habilidades relacionadas à fala nos indivíduos, permitindo o registro de múltiplas avaliações por participante para análises comparativas ao longo do tempo.
Características	Armazena dados adicionais sobre os participantes, incluindo características fenotípicas e comportamentais relacionadas à música e audição.

FONTE: O Autor (2025)

3.7.2 Modelagem banco Alpe

O banco de dados foi modelado com o objetivo de armazenar informações referentes ao estudo de mutações no gene BCHE, localizado na região cromossômica 3q26.1–3q26.2. Os dados foram obtidos através da coleta de amostras sanguíneas compostas por 3.006 doadores de sangue do Centro de Hematologia e Hemoterapia do Paraná. Foram realizados dois estudos para análise e descobrimento de novas mutações no gene.

O primeiro estudo destaca a relevância clínica da enzima BChE, nele buscou-se identificar variantes genéticas no gene BCHE (que codifica a butirilcolinesterase humana). Foram utilizados dois subconjuntos amostrais, extraídos de um total de 3.001 doadores de sangue. As análises foram feitas em regiões específicas do éxon 2, independentemente do fenótipo da BChE, com o intuito de estimar as frequências de novas mutações. O primeiro subconjunto foi composto por 200 doadores de sangue selecionados aleatoriamente, analisados no intervalo de nucleotídeos 155 a 347. O segundo subconjunto incluiu 366 doadores euro-brasileiros do sexo masculino, com fenótipo CHE2 C5-, idade entre 18 e 30 anos e IMC entre 20 e 25, os quais foram analisados nos intervalos de nucleotídeos 155 a 347 e 304 a 535. Como resultado, foram identificadas quatro novas mutações localizadas no éxon 2 do gene BChE: três mutações não sinônimas (223G→C,

resultando em G75R; 270A→C, resultando em E90D; e 297T→G, resultando em I99M) e uma mutação sinônima (486T→C, que codifica a mutação A162A). Todas as variantes apresentaram frequência populacional muito baixa: $0,093 \pm 0,093\%$ para as mutações missense e $0,137 \pm 0,137\%$ para a mutação sinônima (Souza et al., 2005).

O segundo estudo também teve como objetivo identificar e caracterizar mutações naturais no gene BCHE, responsável pela codificação da enzima butirilcolinesterase (BChE), que atua na hidrólise de compostos como succinilcolina, mivacúrio e cocaína. A pesquisa focou em examinar éxons 2 e 4 do gene em uma amostra de 366 doadores de sangue do sul do Brasil, com especial atenção a variantes que poderiam influenciar a atividade enzimática e, conseqüentemente, a resposta a anestésicos. Foram utilizados métodos de triagem molecular, como PCR-SSCA e sequenciamento de DNA, para detectar mutações nos éxons 2 e 4. Mutantes com alterações não sinônimas foram inseridos em vetores de expressão e transfectados em células HEK 293T ou CHO, para produção da enzima recombinante. A atividade da BChE foi medida por ensaios cinéticos com substrato butiriltiocolina, utilizando o método de Ellman. Foram identificadas 12 mutações no gene BCHE, sendo cinco inéditas (K12R, G15G, V294M, G333C e R470W). A variante mais comum foi A539T (variante K), encontrada em cerca de 17,5% dos alelos. A mutação G333C foi a única entre as novas que causou uma redução significativa (80%) na atividade da enzima em relação à forma selvagem. A maioria dos indivíduos com mutações em éxon 2 apresentava baixa atividade de BCHE plasmática, sugerindo uma associação direta entre as variantes e o fenótipo enzimático reduzido (Mikami et al., 2007).

O modelo do banco foi estruturado reaproveitando o modelo do projeto Canta Sabiá, por utilizar a mesma estrutura genética, o SNP relacionado com o Gene. Como complemento, foi adicionada a entidade Alpe que corresponde às amostras dos estudos; nela contém o Índice de Massa Corporal (IMC), idade, altura, peso, etnia, sexo, atividade enzimática do BChE, PN e o RON, que representam a porcentagem de inibição enzimática. Os dados da entidade Alpe foram os resultados das pesquisas citadas anteriormente, relacionados com os SNPs utilizados.

3.7.3 Modelagem banco Projeto Ratos

O banco de dados foi projetado para abrigar múltiplos projetos e experimentos envolvendo ratos, permitindo o armazenamento de dados comportamentais, expressão gênica e atividade enzimática em tecidos específicos dos animais.

As células especializadas dos organismos multicelulares podem alterar a expressão gênica em resposta a sinais extracelulares, como ocorre com as células do fígado quando expostas ao cortisol, que passam a produzir mais proteínas (Alberts, 2017). Assim, a medição da expressão gênica em experimentos realizados antes e depois de uma intervenção, ou em animais submetidos a diferentes tipos de intervenções, pode revelar variações nos níveis de expressão e atividade gênica. Dado modelo visa estruturar esse tipo de experimento a fim de permitir a análise futura desses dados.

A Tabela 4 apresenta a descrição dos objetivos de cada entidade do banco de dados Projeto Ratos. Ademais, os diagramas exibidos nas Seções 3.3 e 3.4 também detalham a estrutura final do modelo.

TABELA 4 – DESCRIÇÃO DOS OBJETIVOS DE CADA ENTIDADE

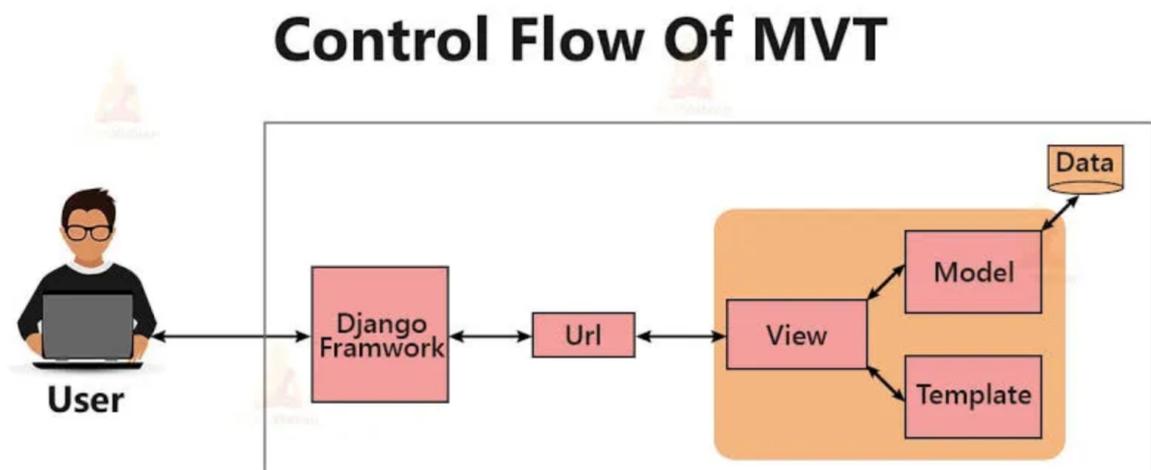
Entidade	Descrição
Projeto	Contém a descrição do projeto e o seu responsável.
Experimento	Guarda os detalhes de cada teste ou estudo realizado. Contém a data do experimento e está relacionada aos tecidos coletados.
Ratos	Armazena dados do animal e o relaciona tanto com o grupo quanto com o tecido.
Grupo	Contém informações sobre o grupo experimental ao qual o rato pertence, estando relacionado à intervenção realizada no projeto. Um rato pode pertencer a mais de um grupo.
Tecido	Contém as principais informações genéticas. Cada amostra possui um identificador único e um tipo. Está ligada ao animal de origem.
Expressão Gênica	Está ligada ao respectivo gene e contém o valor da expressão.
Atividade Enzimática	Relaciona-se aos respectivos genes e contém o valor da atividade e o nome da enzima.

FONTE: O Autor (2025)

3.8 DJANGO

Django é um *framework* web de alto nível para Python, gratuito e de código aberto, que facilita o desenvolvimento rápido e com design limpo. O *framework* é baseado no design *pattern Model-View-Template* (MVT) que trabalha em três blocos principais com o objetivo de interconectá-los para o funcionamento correto da aplicação (Django Software Foundation, 2024). A arquitetura da aplicação pode ser visualizada na Figura 8.

FIGURA 8 – FLUXO DE CONTROLE MODEL VIEW TEMPLATE



FONTE: MAHMOUD, A. (2024)

3.8.1 Model

O Model é o bloco responsável por conter a estrutura lógica do projeto e intermediar dados entre o banco de dados e a *View*. Dentro do arquivo Model são determinados quais tipos

de dados e como serão armazenados dentro do seu banco. Além disso, também é definido como serão exibidos quando forem requisitados pela *View* (Django Software Foundation, 2024). Dentro desse bloco, foram aplicados os modelos relacionais estruturados na seção 3.4.

3.8.2 View

A camada *View* no Django tem como responsabilidade o processamento das requisições do usuário, acesso aos dados necessários por meio da camada *Models* e preparar a apresentação dos dados, feita por meio da renderização de *templates* HTML. A *View* atua como uma ponte entre os dados armazenados e a interface do usuário. No contexto da plataforma desenvolvida, as *Views* foram utilizadas para implementar a lógica de controle da aplicação, como o tratamento das requisições relacionadas ao acesso aos dados genéticos, autenticação de usuários, controle de permissões e manipulação de arquivos para importação e exportação de dados (Django Software Foundation, 2024).

3.8.3 Template

A camada *Template* tem a função de definir como as informações serão apresentadas ao usuário na interface web. Essa camada contém os arquivos HTML que combinam partes estáticas com elementos dinâmicos, que são inseridos com base nos dados processados pela camada *Views*.

O Django utiliza, por padrão, sua própria linguagem de *templates*, conhecida como *Django Template Language (DTL)*. Essa linguagem oferece uma sintaxe simples e expressiva para interpolar variáveis, aplicar filtros e controlar estruturas de repetição e condição, mantendo a separação entre lógica de negócio e apresentação. O mecanismo de *template* do Django funciona em duas etapas: carregamento, o qual ocorre quando o sistema localiza e prepara o *template* para uso, e renderização, que ocorre quando os dados de contexto são inseridos no *template*, gerando o HTML final a ser enviado ao navegador (Django Software Foundation, 2024).

Na plataforma desenvolvida, essa camada apresenta os dados genéticos armazenados no banco de forma organizada e acessível. Foram utilizados blocos reutilizáveis de *layout* e herança de *templates*.

3.9 INTERFACE COM HTML, CSS E INTEGRAÇÃO COM DJANGO

A camada de apresentação da aplicação foi implementada utilizando HTML e CSS. A integração entre Django e HTML ocorreu por meio dos arquivos de *template* (.html), onde foram inseridas expressões e comandos da linguagem DTL do Django. O objetivo da utilização da DTL foi apresentar o conteúdo dinâmico das páginas gerado pelo *back-end*. Ademais, utilizou-se expressões de interpolação “`{{ }}`” para mostrar variáveis enviadas pelas *views*, como listas de registros, nomes de usuários ou mensagens de confirmação. Com essa funcionalidade, é possível ajustar o conteúdo de cada página com base nas informações processadas pelo servidor.

A organização da estrutura da interface foi realizada por meio da herança de *templates*, na qual Bootstrap um arquivo base (base.html) centraliza os elementos compartilhados, possibilitando que as páginas específicas apenas subscrevam os blocos desejados.

Com isso, foi possível manter a identidade visual em diferentes páginas e houve a reutilização de código. O uso do comando “`% block %`” permitiu implementar a herança de *templates* e inclusão de arquivos estáticos, os quais foram utilizados para a reutilização do menu do sistema.

Para melhorar a adaptação visual do aplicativo, foi integrado o *framework Bootstrap*, que possibilita a construção de páginas responsivas e com um layout. Os componentes visuais do *Bootstrap* foram mesclados com classes CSS desenvolvidas.

3.10 CRIAÇÃO DOS BANCOS DE DADOS

Inicialmente, os bancos de dados foram criados no MySQL juntamente com o usuário responsável pelo acesso. As credenciais e configurações de conexão foram devidamente definidas no arquivo *settings.py* do projeto.

Em seguida, a camada de modelos (*models*) do Django foi utilizada para representar e estruturar os dados da aplicação. Cada modelo corresponde a uma tabela no banco de dados, e os atributos definidos nas classes representam as colunas dessas tabelas.

No projeto, os modelos foram definidos nos arquivos *models.py* de cada aplicativo Django, utilizando a classe *models.Model* como base. Por exemplo, o modelo 'Indivíduo' foi criado para armazenar informações demográficas básicas como idade, sexo, escolaridade e área de atuação. Já o modelo 'Característica' foi implementado como uma extensão desses dados, utilizando a relação *OneToOneField*, que permite associar exatamente um conjunto de características a cada indivíduo. Com o intuito de garantir uma organização, os modelos foram separados de acordo com seus respectivos projetos, e conectados a bancos de dados distintos. Essa separação foi feita no arquivo *db_router.py*, que permite o controle dos bancos, podendo escolher o destino de cada aplicação.

3.11 IMPLEMENTAÇÃO DAS PRINCIPAIS FUNCIONALIDADES

Neste tópico, serão apresentados as funcionalidades desenvolvidas para o sistema, bem como a forma como foram implementadas, detalhando as bibliotecas envolvidas.

3.11.1 Implementação das funcionalidades gerenciamento de usuários, controle de grupos e alteração de credenciais de acesso.

Para a implementação do sistema de gerenciamento de usuários e controle de grupos, optou-se por utilizar o *framework Django*, que oferece suporte nativo à autenticação, autorização e gerenciamento de perfis.

O acesso ao gerenciamento de usuário e grupos é feito pelo menu principal da aplicação. O controle de permissões é gerenciado pelo sistema de autenticação do Django, garantindo que apenas usuários com privilégios de superusuário possam acessar tais funcionalidades.

A funcionalidade controle de grupos permite a organização dos usuários de acordo com suas funções na aplicação. Grupos foram definidos com base nos perfis de uso, como administradores, pesquisadores e colaboradores, e receberam permissões específicas de acordo com suas atribuições. Na criação de um grupo, é possível selecionar tabelas e definir, para cada uma delas, o tipo de permissão atribuída: visualização, adição, edição ou exclusão. Podem ser incluídas várias tabelas, inclusive de diferentes bancos de dados, com suas respectivas permissões configuradas individualmente dentro do mesmo grupo.

Para a recuperação de senha via e-mail, foi configurado o SMTP de envio que foi criado no Gmail, permitindo o envio de mensagens para os usuários que solicitassem redefinição de senha. A configuração no arquivo *settings.py* incluiu os parâmetros de acesso ao e-mail. Dessa forma, o sistema envia automaticamente e-mails com links para redefinição de senha, integrando-se às *views* de recuperação do Django, como *PasswordResetView* e *PasswordResetConfirmView*.

3.11.2 Implementação do acesso aos dados

A implementação do acesso e gerenciamento dos dados no sistema foi facilitada pelo uso do módulo administrativo nativo do *framework Django*, configurado por meio do arquivo *admin.py*. Esse módulo oferece uma interface pronta para manipulação dos modelos do banco de dados.

Em cada *model* foi definido um arquivo *admin.py*. Nele, utilizou-se a função *admin.site.register()*, que vincula o modelo à interface administrativa, possibilitando a visualização, criação, edição e exclusão dos dados diretamente pelo painel administrativo do Django.

Além do registro básico, foi realizada a customização das classes de administração, por meio da criação de subclasses de *admin.ModelAdmin*, com o objetivo de melhorar a usabilidade e a eficiência da interface administrativa.

3.11.3 Implementação da importação e exportação dos dados

As funcionalidades de importação e exportação de dados foram integradas ao menu principal do sistema. Os bancos de dados disponibilizados nas funcionalidades de importação e exportação são filtrados com base nas permissões atribuídas ao usuário, de modo que apenas aqueles para os quais o usuário possui acesso autorizado são exibidos.

A funcionalidade de importação de dados foi implementada por meio de uma sequência de três *views* no Django, que organizam o fluxo de entrada do usuário. A primeira função trata a seleção inicial do banco de dados. Caso o formulário seja submetido, a aplicação redireciona para a próxima etapa, enviando o identificador do banco escolhido como parâmetro. Na segunda etapa, o sistema verifica se o banco é válido e exibe as tabelas relacionadas ao banco selecionado. Após a escolha da tabela, a requisição é redirecionada para a função *importar_csv*, que efetivamente realiza a operação de importação. Essa *view* recupera o modelo correspondente à tabela por meio da função *apps.get_model*, utilizando o nome do aplicativo e o nome do modelo definidos no dicionário. O arquivo CSV é então codificado em UTF-8 e processado utilizando *csv.DictReader*, que lê os dados em forma de dicionário. Para cada linha, uma nova instância do modelo é criada com os campos correspondentes e salva no banco de dados indicado. Caso ocorra qualquer erro de correspondência de campos ou falha na validação dos dados, uma mensagem de erro é renderizada informando o problema. O controle de exceções garante que erros específicos durante a importação sejam informados ao usuário. Por fim, ao concluir a operação com sucesso, uma mensagem de confirmação é exibida.

A funcionalidade de exportação de dados foi desenvolvida para a exportação de uma tabela específica ou para a exportação de todas as tabelas de um determinado banco de dados. Essa funcionalidade de exportação de dados foi implementada por meio de uma sequência de três *views* no Django, que organizam o fluxo de entrada do usuário. A primeira função trata a seleção inicial do banco de dados. Caso o formulário seja submetido, a aplicação redireciona para a próxima etapa, enviando o identificador do banco escolhido como parâmetro. Na segunda etapa, o sistema verifica se o banco é válido e exibe as tabelas relacionadas ao banco selecionado. Aqui pode ser escolhido entre exportar todas as tabelas do banco ou escolher uma tabela específica.

Caso o usuário escolha uma tabela, o modelo correspondente à tabela é recuperado a partir dos identificadores do aplicativo e do nome da classe, por meio da API do Django (*apps.get_model*). No caso da exportação unitária, os registros do modelo são extraídos utilizando a *QuerySet* vinculada ao banco selecionado através do método *using()*, e seus campos são serializados em formato CSV. Esse conteúdo é encapsulado em uma resposta HTTP com tipo *text/csv*.

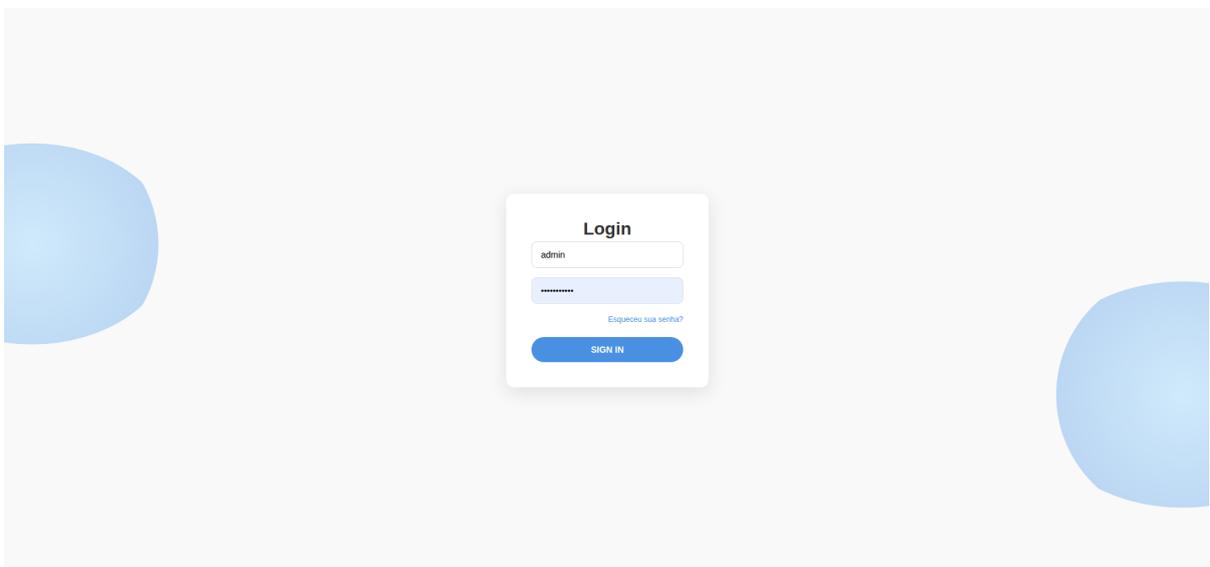
Já na exportação completa, o sistema percorre iterativamente todas as entradas do dicionário de tabelas associadas ao banco, realiza a serialização de dados para cada uma delas, e grava os arquivos gerados em memória utilizando um *buffer* StringIO. Os múltiplos arquivos CSV resultantes são então inseridos em um arquivo compactado .zip por meio da biblioteca *zipfile*, utilizando um *buffer* binário (BytesIO) para evitar escrita em disco. O arquivo final é disponibilizado para o usuário com um cabeçalho adequado para *download* direto pela interface web.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, serão apresentadas as telas implementadas do sistema. No projeto, foi realizada a modelagem e implementação de um banco de dados relacional estruturado a partir de planilhas de pesquisa fornecidas por docentes e pesquisadores da UFPR, bem como de dados provenientes de projetos em andamento. A estrutura criada segue critérios de padronização, garantindo a integridade e a interoperabilidade dos dados genéticos e fenotípicos, conforme os requisitos definidos para o armazenamento e gerenciamento dessas informações.

A Figura 9 representa o resultado da implementação da página de *login* e a Figura 10 a página inicial da plataforma, junto com o seu menu que acessa as demais funcionalidades.

FIGURA 9 – PÁGINA DE LOGIN



FONTE: O Autor (2025)

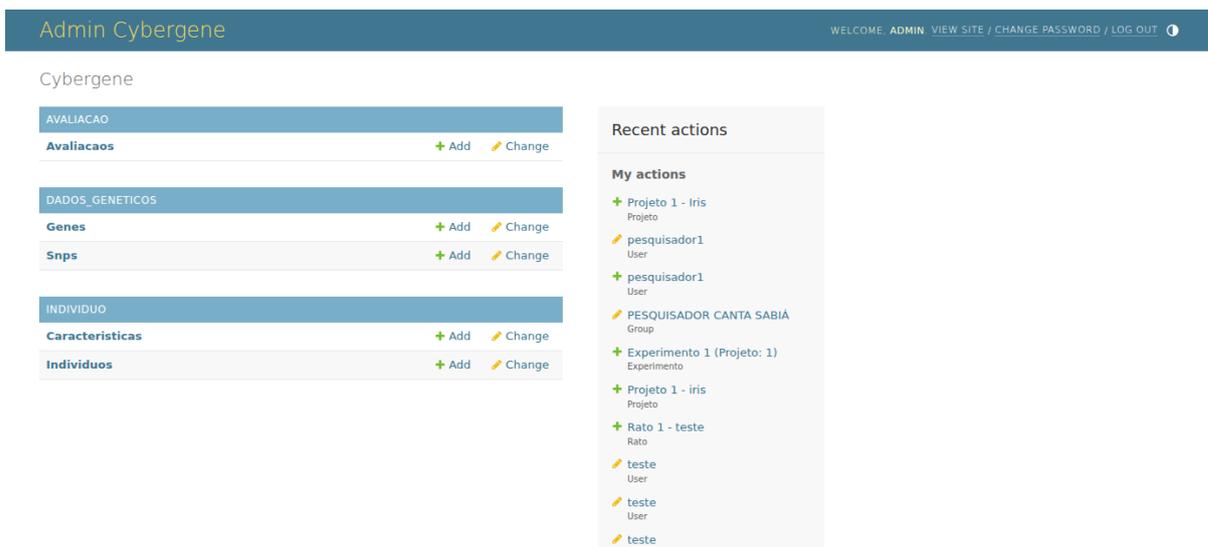
FIGURA 10 – PÁGINA INICIAL



FONTE: O Autor (2025)

Implementaram-se funcionalidades para o registro e a atualização de dados genômicos, incluindo polimorfismos de nucleotídeo único, expressão enzimática, expressão gênica, genes associados e informações diagnósticas. As implementações permitem a inserção de novos registros, bem como a exclusão e modificação dos dados existentes, conforme as necessidades de manutenção e evolução da base de dados. A seguir, a figura 11 ilustra o painel de gerenciamento das informações.

FIGURA 11 – PÁGINA GERENCIAMENTO DAS INFORMAÇÕES



FONTE: O Autor (2025)

Desenvolveram-se mecanismos para a importação de novos dados por meio de planilhas no formato CSV, possibilitando a inserção em lote de informações no banco de dados. Também foi disponibilizada a funcionalidade de exportação dos dados armazenados, permitindo sua extração em formatos abertos para uso em outras ferramentas ou análises externas.

Foi testada a funcionalidade de importação de dados via arquivos CSV padronizados, a qual permite a inserção em massa dos registros. Essa funcionalidade reduz o tempo de trabalho manual e aumenta a eficiência do processo de alimentação da base de dados. Além disso, a exportação de dados em formatos CSV garante a possibilidade de integração com ferramentas estatísticas e de visualização, como R, Python e *softwares* de análise genômica.

Abaixo, as Figuras 12, 13, 14, 15 e 16 ilustram o resultado das telas desenvolvidas para a importação e exportação dos dados.

FIGURA 12 – TELA IMPORTAÇÃO - ESCOLHE BANCO

A imagem mostra a interface de usuário para a importação de dados. À esquerda, há um menu lateral escuro com o título 'Painel' e ícones para: Início, Bancos (com uma seta para baixo), Importação de Dados, Exportação de Dados, Gerenciador de Usuários, Gerenciador de Grupos, Painel admin e Ajuda. O conteúdo principal da tela é branco e apresenta o título 'Escolha a Tabela do Banco: alpe'. Abaixo do título, há um campo de seleção com o texto 'Amostra_alpe' e uma seta para baixo. Abaixo do campo, há um botão verde com o texto 'Avançar'.

FONTE: O Autor (2025)

FIGURA 13 – TELA IMPORTAÇÃO - ESCOLHE TABELA

Painel

- Início
- Bancos
- Importação de Dados
- Exportação de Dados
- Gerenciador de Usuários
- Gerenciador de Grupos
- Painel admin
- Ajuda

Escolha o Banco de Dados para importar

alpe

Avançar

FONTE: O Autor (2025)

FIGURA 14 – TELA IMPORTAÇÃO - ESCOLHE ARQUIVO

Painel

- Início
- Bancos
- Importação de Dados
- Exportação de Dados
- Gerenciador de Usuários
- Gerenciador de Grupos
- Painel admin
- Ajuda

Importar CSV para dados_geneticos_alpe (Banco: alpe)

Selecionar o arquivo CSV

Escolher arquivo Nenhum arquivo escolhido

Importar

FONTE: O Autor (2025)

FIGURA 15 – TELA EXPORTAÇÃO - ESCOLHE BANCO

Painel

- Início
- Bancos
- Importação de Dados
- Exportação de Dados
- Gerenciador de Usuários
- Gerenciador de Grupos
- Painel admin
- Ajuda

Escolha o Banco de Dados

canta_sabia

Avançar

FONTE: O Autor (2025)

FIGURA 16 – TELA EXPORTAÇÃO - ESCOLHE TABELA

Painel

- Início
- Bancos
- Importação de Dados
- Exportação de Dados
- Gerenciador de Usuários
- Gerenciador de Grupos
- Painel admin
- Ajuda

Escolha a Tabela do Banco: canta_sabia

Escolha uma tabela:

SNPs

Exportar CSV

Exportar todas as tabelas (.zip)

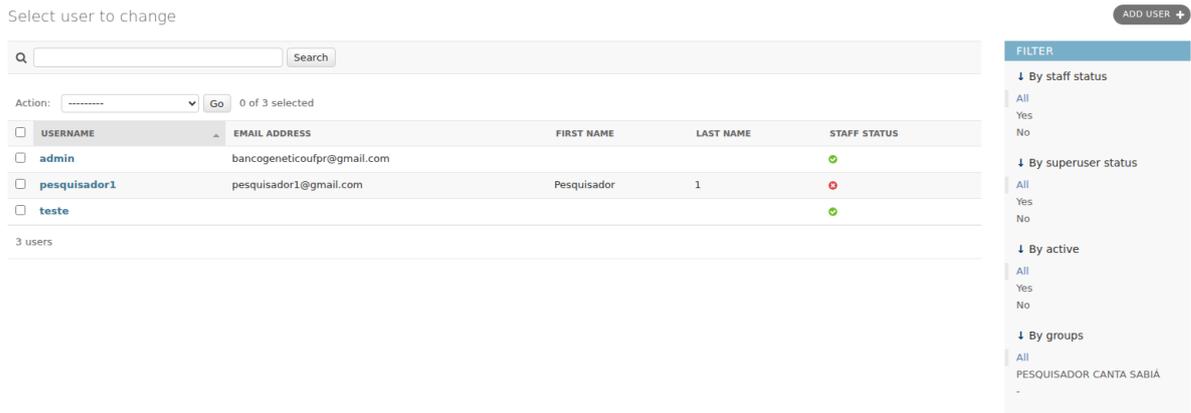
FONTE: O Autor (2025)

Desenvolveu-se um sistema de controle de acesso que permite ao administrador gerenciar os usuários do banco de dados. Foram incluídas funcionalidades para criação, edição e remoção de contas, bem como para a definição de níveis de permissão por meio da criação de grupos de

acesso. Essa estrutura possibilita o controle sobre as operações que cada usuário pode realizar, conforme seu papel no projeto.

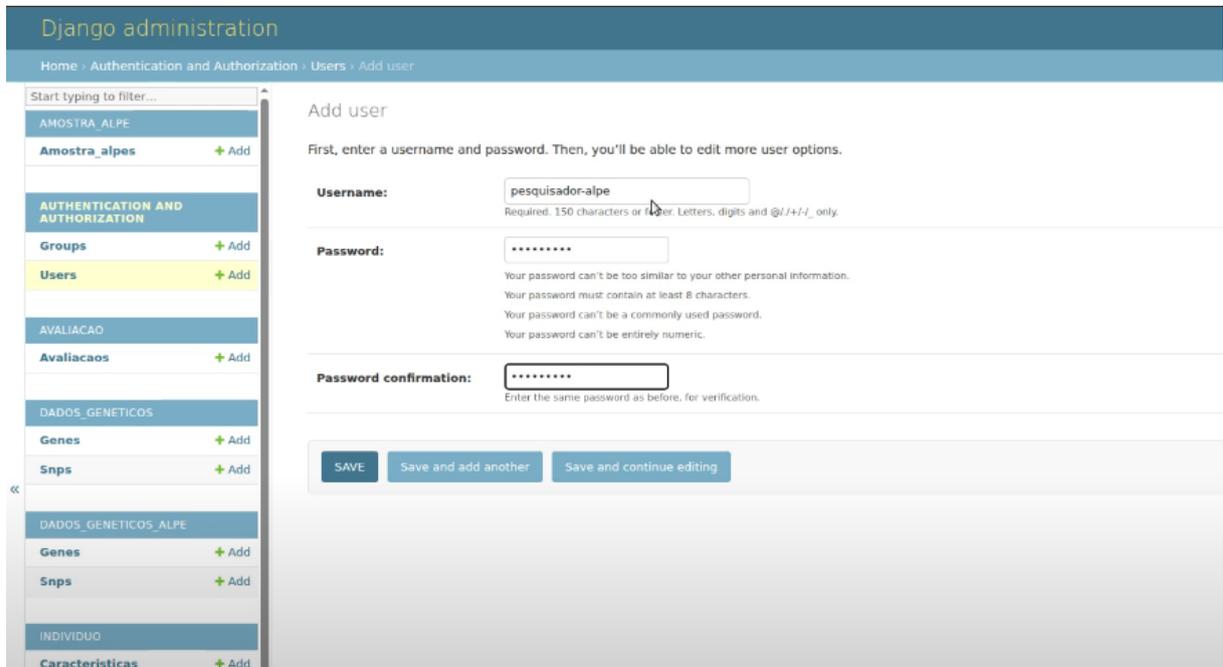
A Figura 17 exibe a tela de gerenciamento de usuários, a Figura 18 exibe a tela de criação de um novo usuário, e a Figura 19 apresenta a interface para criação de grupos com diferentes permissões de acesso ao sistema, permitindo que o administrador configure perfis personalizados conforme o papel do usuário.

FIGURA 17 – TELA GERENCIAMENTO DE USUÁRIO



FONTE: O Autor (2025)

FIGURA 18 – TELA CRIAÇÃO DE USUÁRIO



FONTE: O Autor (2025)

FIGURA 19 – TELA PARA CRIAÇÃO E CONFIGURAÇÃO DE UM GRUPO

Select user to change

ADD USER +

Q Search

Action: Go 0 of 3 selected

<input type="checkbox"/>	USERNAME	EMAIL ADDRESS	FIRST NAME	LAST NAME	STAFF STATUS
<input type="checkbox"/>	admin	bancogeneticoufpr@gmail.com			●
<input type="checkbox"/>	pesquisador1	pesquisador1@gmail.com	Pesquisador	1	●
<input type="checkbox"/>	teste				●

3 users

FILTER

- By staff status
 - All
 - Yes
 - No
- By superuser status
 - All
 - Yes
 - No
- By active
 - All
 - Yes
 - No
- By groups
 - All
 - PESQUISADOR CANTA SABIA
 -

FONTE: O Autor (2025)

Os requisitos do sistema foram definidos e validados por meio de reuniões periódicas com o coordenador deste trabalho. Durante esses encontros, foram discutidas a forma de modelagem dos bancos, bem como as funcionalidades essenciais para a plataforma. Ao final do processo, todos os requisitos definidos inicialmente foram implementados, garantindo o funcionamento adequado da plataforma conforme o planejado.

5 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A plataforma foi desenvolvida com o objetivo de armazenar e estruturar dados de pesquisas conduzidas pelo Departamento de Genética, oferecendo uma base estruturada que facilita a gestão das informações. A integração entre um sistema de acesso adequado e a modelagem de uma estrutura coerente com as demandas das pesquisas foi fundamental para a construção da plataforma.

A plataforma tem capacidade de integrar dados genéticos, fenotípicos e ambientais, garantindo integridade, padronização e interoperabilidade das informações. A modelagem do banco de dados é baseada em princípios relacionais.

Por se tratar de um trabalho multidisciplinar, foi essencial a troca de informações com o coorientador, por meio de reuniões frequentes e revisões do material, alinhando a modelagem dos bancos aos objetivos científicos das pesquisas envolvidas.

Todos os requisitos inicialmente propostos foram implementados. A criação da plataforma atendeu à necessidade de padronizar e consolidar informações que, anteriormente, encontravam-se dispersas em diferentes formatos e fontes, dificultando a realização de análises integradas e a continuidade das pesquisas.

A partir disso e de ideias que surgiram durante o desenvolvimento, foi possível formular algumas recomendações futuras, sendo elas:

1. Possibilidade de integrar os dados genéticos com outros bancos de dados, com o objetivo de auxiliar na análise dos dados. O Ensembl, por exemplo, disponibiliza uma série de recursos de código aberto para acesso a seus bancos de dados genômicos (ENSEMBL, 2024).
2. A integração de um formulário diretamente no sistema torna o processo de coleta mais ágil, organizado e acessível. Os formulários são ferramentas amplamente utilizadas para a coleta de dados individuais.
3. A plataforma poderá incorporar um sistema de notificação que informe, em tempo real, os responsáveis pelo banco de dados sobre os acessos realizados pelos usuários. Essa funcionalidade visa aumentar o controle e a segurança do sistema, permitindo o monitoramento da atividade dos usuários.
4. A aplicação de testes com usuários reais, como professores e alunos. Esse processo permite avaliar a usabilidade, funcionalidade e eficiência do sistema em contextos práticos, além de identificar possíveis dificuldades, limitações e oportunidades de melhoria com base na experiência dos próprios usuários.
5. Permitir que pesquisadores possam criar projetos de forma automatizada, utilizando dados genéticos e fenotípicos de acordo com suas necessidades específicas. Essa funcionalidade possibilita ao usuário selecionar as entidades e atributos que deseja incluir, personalizando a estrutura do banco conforme os objetivos da pesquisa, sem a necessidade de conhecimentos técnicos avançados em modelagem de dados.

REFERÊNCIAS

- Alberts, B. (2017). *Fundamentos da Biologia Celular*. ArtMed, Porto Alegre, 4 edition. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788582714065/>. Acesso em: 29 jun. 2025.
- ALVES, W. P. (2014). *Banco de Dados*. Editora Érica. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788536518961/>. Acesso em: 09 jun. 2025.
- Antonio, B. S. M., Souza, R. L. R., Quadrado, S. E., Lehtonen-Souza, L. M. F. O. e Alle, L. F. (2024). Genes, language, and musicality: An intersection. *Ciências da Saúde*, 28(135). Publicado em: 05 jun. 2024. Disponível em <https://revistaft.com.br/genes-language-and-musicality-an-intersection/>. Acesso em 07 abr. 2025.
- Cardoso, G. C. e Cardoso, V. M. (2012). *Sistemas de Banco de Dados*. Saraiva, Rio de Janeiro, 1ª edition. E-book, p. 107. ISBN 9788502162839. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788502162839/>. Acesso em: 22 jun. 2025.
- Django Software Foundation (2024). Transactions | django documentation. <https://docs.djangoproject.com/en/5.0/topics/db/transactions/>. Cobre transaction.atomic() e controle de concorrência. Acesso em: 30 jun. 2025.
- Elmasri, R. e Navathe, S. (2005). *Sistemas de Banco de Dados: Fundamentos e Aplicações*. Pearson, São Paulo, 4 edition.
- ENSEMBL (2024). Ensembl Genome Browser. <https://www.ensembl.org>. Acessado em 01/07/2025.
- Machado, F. N. R. (2024). *Banco de Dados – Projeto e Implementação*. Érica, Rio de Janeiro, 4 edition. E-book, p. 10. ISBN 9788536532691. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788536532691/>. Acesso em: 29 jun. 2025.
- MedlinePlus (2025). Polimorfismos de nucleotídeo único (snps). <https://medlineplus.gov/genetics/understanding/genomicresearch/snp/>. Acesso em: 29 jun. 2025.
- Mikami, L. R., Wieseler, S., Souza, R. L., Schopfer, L. M., Nachon, F., Lockridge, O. e Chautard-Freire-Maia, E. A. (2007). Five new naturally occurring mutations of the bche gene and frequencies of 12 butyrylcholinesterase alleles in a brazilian population. *Pharmacogenetics and Genomics*, 17(3):231–238.
- NCBI RefSeq (2010). Foxp2 forkhead box p2 [homo sapiens (human)]. <https://www.ncbi.nlm.nih.gov/gene/93986>. “This gene is required for proper development of speech and language regions of the brain during embryogenesis. . .” Acesso em: 30 jun. 2025.

- Pressman, R. S. e Maxim, B. R. (2020). *Engenharia de Software: Uma Abordagem Profissional*. McGraw-Hill Brasil, São Paulo, 8 edition. Disponível em <https://integrada.minhabiblioteca.com.br/reader/books/9786558040118/>. Acesso em 04 mai. 2025.
- Silberschatz, A., Korth, H. F. e Sudarshan, S. (2020). *Sistemas de Banco de Dados*. Bookman, 7 edition.
- Sommerville, I. (2011). *Engenharia de Software*. Addison-Wesley, São Paulo, 9 edition.
- Souza, R. L., Mikami, L. R., Maegawa, R. O. e Chautard-Freire-Maia, E. A. (2005). Four new mutations in the bche gene of human butyrylcholinesterase in a brazilian blood donor sample. *Clinica Chimica Acta*, 357(2):153–160. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1096719204003257?via>
- Strachan, T. e Read, A. (2013). *Genética Molecular Humana*. ArtMed, Porto Alegre, 4 edition. E-book, p. 229. ISBN 9788565852593. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788565852593/>. Acesso em: 29 jun. 2025.
- Sunyaev, S. R., Lathe, W. C., Ramensky, V. E. e Bork, P. (2000). Snp frequencies in human genes: an excess of rare alleles and differing modes of selection. *Trends in Genetics*, 16(8):335–337. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/10904261/>. Acesso em: 14 mai. 2025.
- The International SNP Map Working Group (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822):928–933. Disponível em: <https://www.nature.com/articles/35057149>. Acesso em: 05 jun. 2025.
- Thompson, R. L. e Thompson, A. D. (2016). *Thompson & Thompson: Genética Médica*. Elsevier. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788595151819/>. Acesso em: 09 jun. 2025.
- Turnpenny, P. D. e Ellard, S. (2021). *Emery's Elements of Medical Genetics and Genomics*. Elsevier, Philadelphia, 16 edition. E-book. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788595151819/>. Acesso em: 09 jun. 2025.