



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de Bauru

Adriano Requena

Ferramenta de Visualização para Exibição de Dados sobre Genética de Peixes

Bauru, São Paulo, Brasil

Maio 2023

Adriano Requena

Ferramenta de Visualização para Exibição de Dados sobre Genética de Peixes

Dissertação de Mestrado elaborada junto ao Programa de Pós-Graduação em Ciência da Computação – Área de Concentração de Computação Aplicada, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação

Universidade Estadual Paulista "Júlio de Mesquita Filho"

Faculdade de Ciências

Programa de Pós-Graduação em Ciência da Computação

Orientador: Dr. José Remo Ferreira Brega

Bauru, São Paulo, Brasil

Maior 2023

R427f Requena, Adriano
 Ferramenta de Visualização de Dados para Exibição de Dados
 sobre Genética de Peixes / Adriano Requena. -- Bauru, 2023
 80 p. : tabs., fotos

 Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
 Faculdade de Ciências, Bauru
 Orientador: José Remo Ferreira Brega

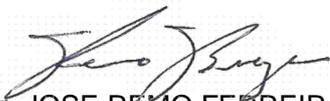
 1. Ciência da Computação. 2. Visualização da Informação. 3.
 Genética de Peixes. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências, Bauru. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE ADRIANO REQUENA, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO, DA FACULDADE DE CIÊNCIAS - CÂMPUS DE BAURU.

Aos 25 dias do mês de maio do ano de 2023, às 14:00 horas, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de ADRIANO REQUENA, intitulada **Ferramenta de Visualização para Exibição de Dados sobre Genética de Peixes**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. JOSE REMO FERREIRA BREGA (Orientador(a) - Participação Virtual) do(a) Departamento de Computacao / UNESPCampus de Bauru, Prof. Dr. KELTON AUGUSTO PONTARA DA COSTA (Participação Virtual) do(a) FC / UNESP Bauru SP, Profa. Dra. NATASHA MALVEIRA COSTA VALENTIM (Participação Virtual) do(a) Departamento de Informática / Universidade Federal do Paraná. Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final APROVADO. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.



Prof. Dr. JOSE REMO FERREIRA BREGA

Agradecimentos

Expresso aqui em especial a minha eterna gratidão ao meu orientador, Prof. Dr. José Remo Ferreira Brega, ou simplesmente "Remo", como todos o tratam, me acolheu como um filho e propiciou a oportunidade do estudo e evolução, aprendi muito com a sua orientação, sempre solicito mesmo em tempos de pandemia. Expresso, ainda, minha gratidão, a todos os professores que tive oportunidade de conhecer e que me transmitiram conhecimentos em suas disciplinas, bem como, a todos os colegas do mestrado que puderam contribuir no aprendizado especialmente ao Luiz Felipe de Camargo e à Juliana Feitosa que sempre me incentivaram e auxiliaram durante o mestrado. Agradeço também à minha família, principalmente, minha esposa Eloá que sempre me incentivou, foi meu porto seguro e, por vezes, até me orientou, e aos meus filhos Isa e Leo, que apesar de serem crianças, sempre entenderam o meu envolvimento com os estudos. Agradeço à Soft-Line, empresa na qual trabalho pela oportunidade do estudo, e, principalmente a Deus, pela oportunidade da vida, e pela evolução como ser humano, não posso esquecer do meu amigo Fabrício Quintanilha Baptista que também fez parte do programa de pós me incentivou a entrar para o mestrado e Ricardo Mazzarioli com a sua paciência me orientou no desenvolvimento da ferramenta. E por fim, ao Programa de Pós-Graduação em Ciência da Computação (PPGCC) da Unesp, por todo apoio e pela qualidade do programa.

Resumo

O uso de técnicas de visualização da informação e as pesquisas sobre genética têm melhorado muito o conhecimento na medicina, na saúde, na reprodução de animais e até no próprio comércio. Geralmente os dados são de estruturas complexas e de difícil compreensão. Os pesquisadores têm obtido padrões e conhecimento sobre esses dados, e, com isso, os processos de descoberta sobre medicamentos para doenças, melhoria no cultivo de culturas e na própria criação de animais para um melhor rendimento em cada uma dessas áreas. O objetivo é apresentar uma ferramenta que utiliza técnicas de Visualização da Informação que será utilizada no acompanhamento da melhoria genética na reprodução de peixes da Unesp de Jaboticabal. O levantamento dos requisitos foi efetuado com doutores e estudantes do Doutorado da Área de Aquicultura da Unesp da cidade de Jaboticabal, Estado de São Paulo. Essas contribuições foram muito importantes para entendimento do processo para aplicação da ferramenta de Visualização da Informação. A contribuição da ferramenta para os pesquisadores está na seleção das melhores progenies que são fruto dos acasalamentos dos peixes da mesma espécie, nessa pesquisa trabalhamos com o Pacu, Tambaqui e Tilápia.

Palavras-chave: Visualização Informação, Genética, Técnicas de Visualização.

Abstract

The use of information visualization techniques and research on genetics have greatly improved knowledge in medicine, health, animal reproduction and even commerce itself. Generally, the data have complex structures and are difficult to understand. Researchers have obtained results and knowledge about this data, and, with it, the processes of discovery about medicines for diseases, improvement in the cultivation of cultures and in the creation of animals for a better yield in each of these areas. The objective is to present a tool that uses Information Visualization techniques that will be used in the monitoring of the genetic improvement in the reproduction of fish from Unesp de Jaboticabal. The survey of requirements was approved with doctors and PhD students from the Aquaculture Area of Unesp in the city of Jaboticabal, State of São Paulo. These contributions were very important for understanding the process for applying the Information Visualization tool. The contribution of the tool to the researchers is in the selection of the best progenies that are the result of the mating of fish of the same species, in this guaranteed search with Pacu, Tambaqui and Tilápia.

Keywords: Visualization Information, Genetics, Visualization Techniques.

Lista de ilustrações

Figura 1 – Classificação final dos estudos	29
Figura 2 – Evolução da quantidade de estudos	29
Figura 3 – Foto do Pacu com as referências para medidas	36
Figura 4 – Imagem da Planilha de Fenótipos	37
Figura 5 – Figura da Ferramenta de Análise da Funcionalidade de Importação	41
Figura 6 – Figura da Lista de Peixes	42
Figura 7 – Dashboard	44
Figura 8 – Figura da Técnica de Visualização de Diagrama de Cordas	45
Figura 9 – Figura da Técnica de Visualização de Coordenadas Paralelas	46
Figura 10 – Figura da Técnica de Visualização de Diagrama de Acordes com a Legenda	47
Figura 11 – Figura da Técnica de Visualização de Coordenadas Paralelas com a seleção de Linhas	48
Figura 12 – Resultado sobre a Idade dos Pesquisadores	50
Figura 13 – Resultado Sobre o Genêro dos Pesquisadores	51
Figura 14 – Resultado Sobre o Nível de formação dos Pesquisadores	51
Figura 15 – Resultado sobre o Estado Atual da Formação dos Pesquisadores	51
Figura 16 – Resultado sobre a Área de Formação dos Pesquisadores	52
Figura 17 – Resultado sobre a Interface	52
Figura 18 – Resultado sobre a Interface da Lista de Peixes	53
Figura 19 – Resultado da Usabilidade do Filtros	53
Figura 20 – Resultado da Usabilidade da Ferramenta	54
Figura 21 – Resultado da Análise da Seleção das Linhas da Coordenadas Paralelas	55
Figura 22 – Resultado da Análise das Cores Utilizadas no Gráfico de Cordas	56
Figura 23 – Resultado da Análise da Utilização da Legenda	57
Figura 24 – Resultado da Análise sobre as Características	57
Figura 25 – Resultado da Análise sobre as Medidas e os EBV	58
Figura 26 – Resultado da Análise para Identificação de Padrões	58
Figura 27 – Resultado da Análise das Técnicas de Visualização Utilizadas	59
Figura 28 – Resultado da Análise das Características na Área de Pesquisa Resistência	59
Figura 29 – Resultado da Análise das Características na Área de Pesquisa Morfometria	60
Figura 30 – Resultado sobre a Análise Geral	60
Figura 31 – Resultado da Análise da Correlação entre as Características	61

Lista de tabelas

Tabela 1 – Exemplos de Gráficos e suas aplicações	21
Tabela 2 – Resultados — Fase de Identificação	27
Tabela 3 – Resultados — Fase de Identificação — Após alterações	28
Tabela 4 – Quantidades de estudos durante as fases da revisão	28
Tabela 5 – Ocorrência de tipos de estudo sobre genética	30
Tabela 6 – Técnicas visuais mais utilizadas	31
Tabela 7 – Fontes de dados utilizadas	32
Tabela 8 – Linguagens de programação	32
Tabela 9 – Artigos incluídos na revisão.	71
Tabela 10 – Artigos excluídos na revisão.	79

Lista de abreviaturas e siglas

HTML5	Hypertext Markup Language versão 5
StArt	State of the Art through Systematic
GPS	Global Positioning System
CSV	Comma-Separated-Values
EBV	Estimated Breeding Values

Sumário

1	INTRODUÇÃO	13
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	Genética	15
2.2	Genética no Estudo de Doenças	16
2.3	Genética na Reprodução de Animais	16
2.4	Genética na Reprodução de Peixes	17
2.5	Visualização da Informação	17
2.5.1	Definindo a Visualização de Dados	17
2.5.2	Os oito atores envolvidos no Projeto de Visualização	18
2.5.3	Técnicas	18
2.6	Combinando Técnicas de Visualização	22
3	METODOLOGIA	23
3.1	Revisão Sistemática da Literatura	23
3.2	Objetivos da Revisão	23
3.3	Questões de Pesquisa	24
3.4	Criação do protocolo	24
3.4.1	Estratégia de busca	24
3.4.2	Bases de dados	25
3.4.3	Critérios de inclusão e exclusão	25
3.4.3.1	Critérios de inclusão	25
3.4.3.2	Critérios de exclusão	25
3.4.4	Extração de Dados	26
3.4.5	Síntese dos estudos	27
3.5	Ferramenta StArt	27
3.6	Condução	27
3.7	Análise e discussão dos resultados	28
3.8	Estudos	28
3.8.1	Estudos excluídos	29
3.8.2	Estudos Não Selecionados	30
3.8.3	Estudos incluídos	30
3.9	Tipos de Estudos sobre Genética	30
3.10	Técnicas de Visualização	30
3.11	Objetivos da Visualização	31
3.12	Formato de dados	32

3.13	Desenvolvimento e linguagem	32
3.14	Relevância do estudo	33
3.15	Limitações do estudo	33
3.16	Tendências	33
3.17	Considerações Finais	34
4	FERRAMENTA DE VISUALIZAÇÃO GENÉTICA DE PEIXES	35
4.1	Cenário Atual	35
4.2	Tecnologias Envolvidas	38
4.2.1	CSharp	38
4.2.2	Arquitetura .Net	38
4.2.3	Java Script	38
4.2.4	SQLite	39
4.2.5	D3	39
5	DESENVOLVIMENTO DA FERRAMENTA	40
5.1	Módulos	40
5.2	Importador	40
5.3	Lista de Peixes	41
5.4	Dashboard	42
5.5	Diagrama de Cordas	44
5.6	Coordenadas Paralelas	45
5.7	Utilização da Ferramenta	46
6	AVALIAÇÃO	49
6.1	Processos	49
6.2	Questionário	49
6.3	Video	50
6.4	Questionário e Respostas	50
6.5	Identificação	50
6.6	Interface	52
6.7	Usabilidade	53
6.8	Análise	55
7	CONCLUSÃO	62
7.1	Observações Finais	62
7.2	Contribuições do Trabalho	62
7.3	Trabalhos Futuros	62
7.4	Participações em Workshop e Publicações	63

	REFERÊNCIAS	64
A	DADOS DOS ESTUDOS INCLUÍDOS	70
B	DADOS DOS ESTUDOS EXCLUÍDOS	79

1 Introdução

O avanço na pesquisa e tecnologia sobre genética melhorou significativamente nossa compreensão da biologia, saúde e medicina. Os dados sobre esse assunto são muito complexos e contêm informações de genótipo e fenótipo. Há muito tempo, pesquisadores de saúde sabem que muitas doenças, como câncer são hereditários. Obter insights e compreensão desses dados permitiria uma melhor compreensão da correlação entre genes e doenças, o que poderia facilitar a personalização de tratamento para os pacientes (Khwaja; Kalofonou; Toumazou, 2017).

A genética está associada a várias áreas, como reprodução de animais, atividades esportivas, e descoberta de medicamentos para combater doenças. O estudo da própria doença pode possuir alguma relação com a genética. Na medicina, o diagnóstico diferencial clínico de doenças hereditárias determina a doença que aflige um paciente, e é um processo complexo e iterativo. Esta prática envolve um esforço na coleta significativa dos dados, incluindo a revisão do histórico médico do paciente e da família (Glueck et al., 2016a).

Na reprodução de animais a genética possui grande relevância na descoberta do conjunto de informações que podem auxiliar no nascimento de um animal melhor geneticamente. Avanços na microscopia de fluorescência, design de fluorescente sondas, métodos experimentais e linhagens animais transgênicas pavimentaram o caminho para a biologia quantitativa baseada em imagens. Com a capacidade de criar imagens de espécimes inteiros em nível celular com alta resolução espacial e temporal e para adquirir amostras múltiplas com configurações de alto rendimento, somos capazes de obter percepções sem precedentes sobre os processos de desenvolvimento e sua variabilidade no nível da população (Thierbach et al., 2016a).

O objetivo principal desta dissertação é apresentar os resultados obtidos através da realização de uma revisão sistemática que contribuiu para o entendimento sobre Técnicas de Visualização utilizadas em dados de genética. Esta pesquisa serviu de base para aumentar o conhecimento sobre assunto e resultou no desenvolvimento da ferramenta que possibilite o melhor entendimento dos pesquisadores sobre a genética de peixe utilizando técnicas de Visualização da Informação. Nessa revisão, serão apresentados como essas ferramentas foram desenvolvidas, qual a linguagem de programação utilizada na sua construção, qual o formato de dados que elas utilizam para a exibição da informação e o mais importante, as técnicas de Visualização que foram empregadas sobre esses dados para a análise das pessoas envolvidas no processo. Com base na pesquisa efetuada e no conhecimento obtido, as informações foram aplicadas no desenvolvimento de uma ferramenta que utiliza técnicas de Visualização da Informação baseadas nos dados dos peixes resultando num dashboard.

O Capítulo 2 deste trabalho trata dos conceitos e técnicas utilizadas, seguido do Capítulo 3 que descreve a metodologia, resultados, discussão e considerações finais da revisão sistemática da literatura. Depois o capítulo 4 que aborda o cenário, técnicas e tecnologias para solucionar o problema, no capítulo 5 aborda os módulos e opções da ferramenta desenvolvida, no capítulo 6 é demonstrado o resultado obtido através do questionamento efetuado aos pesquisadores e por fim o último capítulo 7 que descreve a conclusão desse trabalho.

2 Fundamentação Teórica

Neste Capítulo são apresentados conceitos para direcionar a presente revisão. Dentre eles estão conceitos que buscam definir as áreas envolvidas, a Genética e a Visualização da Informação. Para a área de genética são definidos também os conceitos relacionados ao estudo de doenças, a genética aplicada na reprodução de animais. Posteriormente é apresentado o conceito de Visualização da Informação.

2.1 Genética

A genética é a parte da ciência que estuda a hereditariedade, a estrutura e função dos genes e a variação dos seres vivos. É através da genética que buscamos compreender os mecanismos e leis de transmissão das características através das gerações (GRIFFITHS A. J. F., 2009). Os primeiros estudos da genética na ciência moderna iniciaram-se com o monge austríaco Gregor Mendel, na década de 1860. Realizando cruzamentos entre linhagens de ervilhas, ele observou a existência de “fatores” distintos que eram transmitidos dos genitores para a prole.

Os padrões de herança observados por Mendel correspondem aos padrões de distribuição dos cromossomos nos gametas no processo da meiose. Mais tarde, com o aprofundamento das pesquisas, os “fatores” foram chamados de genes. Antes de Mendel, um nobre húngaro, Imre Festetics, foi o primeiro a usar o termo genética. Em seu livro, ele descreveu várias regras sobre herança genética, tendo sido o primeiro a descrever os princípios básicos da mutação.

Em genética, a transmissão de informação envolve quatro elementos básicos: a diversidade de estruturas celulares com inúmeras formas especificando aspectos diferentes do organismo; um mecanismo de replicação que copia a informação e repassa à prole; a capacidade de mutação para que haja maior variabilidade; e a capacidade da prole em traduzir as informações herdadas em proteínas. O método genético fundamental para o estudo dos fenômenos biológicos é a diferença genética entre os indivíduos ou variação genética. Essa variabilidade pode ser de ocorrência natural entre os indivíduos da mesma espécie, caracterizando o polimorfismo genético. Mas pode também ser causada por mutações, onde cada gene mutante revelará um componente genético do processo.

2.2 Genética no Estudo de Doenças

Em epidemiologia genética, estudos de associação ampla do genoma examinam muitas variantes genéticas em diferentes indivíduos para verificar se alguma variante está associada a um traço específico, como uma doença. A genotipagem do polimorfismo de nucleotídeo único é usada para identificar tais variantes. Um dos muitos desafios dos estudos depende das amostras representativas de um subconjunto de indivíduos de uma população. Informações como a frequência de alelos na população são estimadas a partir dessas amostras representativas. O problema é que essas estimativas estão sujeitas a inflação ou deflação quando os dados da genotipagem são derivado de indivíduos com famílias não relatadas ou relacionamentos ou com ancestralidade mista, potencialmente levando à estratificação indesejada da população. Portanto, é essencial que as amostras sejam compostas de indivíduos com o mais fraco relacionamento familiar ou ancestralidade mista.(Aupetit et al., 2016).

As taxas de cura do câncer melhoraram drasticamente, graças ao crescimento das tecnologias e da pesquisa. Ainda uma das prioridades da pesquisa é encontrar novas maneiras de combater o câncer com descobertas de ponta que são mais seguras e mais eficazes. Tratamentos usando métodos clínicos isolados podem não ser suficientes conseguí-lo. O genoma humano é complexo, e com milhares de genes contendo informações sobre os pacientes e os mecanismos biológicos de suas doenças. Combinar informações genômicas com dados clínicos e capturar a complexidade do genoma de maneira significativa e acionável, o atendimento personalizado poderia adaptar os tratamentos médicos para atender às necessidades exclusivas dos indivíduos(Aupetit et al., 2016).

2.3 Genética na Reprodução de Animais

Uma questão importante para os biólogos e ornitologistas, em particular, é a exploração de dados a respeito de suas espécies de estudo, especialmente se estas forem reunidas em muitos anos e estivem ligadas a outras fontes de dados. Na verdade, nas últimas décadas, com o desenvolvimento de novas tecnologias para registrarem os movimentos dos animais (por exemplo, GPS e acelerômetros), novos sensores para medidas ambientais e fisiológicas aumentaram a quantidade de dados disponíveis para pesquisadores de campo aumentaram dramaticamente e as pesquisas se deparam com interpretação e análise de grandes conjuntos de dados (Palleschi; Crielesi, 2019).

Novos paradigmas para estudos de biodiversidade foram recentemente propostos por (KELLING W. M. HOCHACHKA; HOOKER, 2009). Para lidar com grandes volumes de dados, uma nova disciplina chamada "Ciência Intensiva de Dados" com base em dados de abordagem dirigida foi sugerida para explorar a complexidade de sistemas ecológicos, especialmente quando se considera grandes escalas espaciais e temporais. Visualização de

dados com uma variedade de novas abordagens, oferece uma ferramenta de exploração que pode ajudar a descobrir padrões complexos contidos nos dados e inspirar novas hipóteses e análises.

2.4 Genética na Reprodução de Peixes

A aquicultura continental de espécies nativas na América do Sul ainda é resultante de peixes não selecionados geneticamente, revelando a necessidade de projetos de melhoramento genético para espécies neotropicais, particularmente no pacu. O desempenho de crescimento é o principal objetivo de reprodução para aquicultura espécies (GJEDREM T., 2018). Atualmente, não há informações sobre parâmetros genéticos para características de crescimento em pacu. Em peixes, a alta número de estudos relacionados ao crescimento reflete a importância econômica de aumentar o peso final mais rapidamente. Como é mais fácil de medir, muitos programas de melhoramento preferem selecionar apenas o peso como critério de seleção; no entanto, estudos associando peso corporal e comprimento ganharam destaque na aquicultura, o que sugere que a seleção para características é viável para espécies de aquicultura, especialmente porque pode resultar em melhor rendimento de filé (KAUSE A., 2013).

2.5 Visualização da Informação

Os conceitos dessa seção definem a Visualização da Informação, os atores envolvidos e as técnicas de um projeto de visualização. (KIRK, 2012a).

2.5.1 Definindo a Visualização de Dados

A Visualização da Informação é um campo dinâmico e em evolução, e esta singular conjunção de arte e ciência para moldar sua prática, significa que uma única definição perfeita, e universalmente acordada vai ser sempre difícil de construir. A metodologia irá introduzir uma série de pensamentos e decisões-chave em cada fase do processo. Estes objetivos transcendem qualquer passo individual e destacam as questões complexas que você tem que lidar com todo o seu processo. Logo a seguir serão explicados os objetivos e os oito atores envolvidos no projeto de visualização.

Os principais objetivos do Projeto de Visualização :

- Lutar pela forma e função;
- Justificar a seleção de tudo o que fazemos;
- Criar acessibilidade por meio de um design intuitivo; e

- Nunca enganar o receptor;

2.5.2 Os oito atores envolvidos no Projeto de Visualização

- O iniciador: é o líder, a pessoa que está à procura de uma solução para a tarefa, vai estabelecer o funcional e o tom da direção do projeto, bem como identificar e traçar o perfil do público-alvo e responsável pela escolha da plataforma;
- O cientista de dados: é caracterizado como o mineiro de dados, e é responsável pelo fornecimento, aquisição, manuseio e preparação dos dados;
- O jornalista: é o contador de histórias, a pessoa que estabelece a abordagem narrativa para o contexto do problema de visualização e trabalhando com o cientista de dados e o iniciador, eles são capazes de estabelecer as histórias principais e ângulos com que a fatia da análise;
- O cientista da computação: é o executor, quem produz o projeto vivo com sua capacidade técnica crítica, ele é quem irá construir a solução, e também reforçará o cientista de dados com seus conhecimentos técnicos para lidar de forma mais eficiente com a coleta de dados, manipulação e atividades de visualização de préprodução;
- O designer: é o criativo que em harmonia com o cientista da computação vai entregar a solução;
- O cientista cognitivo: é o pensador, em termos de apreciar a ciência por trás da eficácia das soluções técnicas e personalizadas, e possui a percepção de conhecimento sobre como o olho e o cérebro trabalham de forma mais eficaz;
- O comunicador: tem a função de ser um facilitador na comunicação dos envolvidos no projeto; e
- O gerente de projeto: este último papel é essencialmente o de gerente ou coordenador, a pessoa que faz muito para pegar muitas das tarefas impopulares para ajudar a trazer todo o projeto em conjunto.;

2.5.3 Técnicas

Com o advento do Big Data, o volume de dados se tornou algo extremamente abundante, e, para conseguirmos extrair informações dos dados, as técnicas de visualização têm um papel muito importante nessa fase, auxiliando na extração de informação do dado em si agregando tendências, direção e atingindo seu objetivo que é transformar um dado em informação. Para a escolha de um método correto de visualização de dados, existe um certo número de formas de classificação, e uma variedade de métodos, mas, aqui, há uma taxonomia sugerida:

- Comparando os valores categóricos: Para facilitar as comparações entre tamanhos relativos e absolutos de valores categóricos;
- Avaliando hierarquias e relacionamentos parte-de-um-todo: Para fornecer uma repartição de valores categóricos na sua relação a uma população de valores ou como constituinte de elementos de estruturas hierárquicas;
- Mostrando mudanças ao longo do tempo: Para explorar dados temporais e mostrar as novas tendências e os padrões de valores ao longo de um período de tempo contínuo;
- Mapeando dados geoespaciais: Para traçar e conjuntos de dados atuais com propriedades geo-espaciais através das muitas estruturas diferentes de mapeamento;
- Relacionamentos de gráficos e de representação gráfica:

Quando se fala em um gráfico, o de área empilhada mostra alterações ao longo do tempo, mas também facilita a comparação categórica das suas diferentes camadas. Depois de identificado o método geral de visualização, começa-se a filtrar ainda mais para baixo, identificando-se os tipos de gráficos mais adequados.

Outro fator importante é a determinação do grau de precisão da interpretação do usuário. A representação visual atingiu o objetivo na demonstração dos dados de forma a passar a informação esperada? Na exibição dos dados existem alguns fatores que possibilitam a melhora da representação dos dados na visualização. O uso da cor é uma importante potencial da variável visual para representação dos dados. "Cores bem utilizadas podem melhorar e clarificar uma apresentação. Cores mal usadas mal vão obscurecer, atrapalhar, e confundir." apud citada por Maureen Stone (KIRK, 2012a).

A interatividade é um outro fator que traz um diferencial para a visualização, ela sendo estática é uma poderosa fotografia concebida com cuidado, organizadas e executa a visão que consegue retratar a sequência ou movimento de uma história sem a implantação real do movimento. As melhorias na tecnologia ao longo da última década criaram oportunidades incríveis para desenvolvedores talentosos para construir visualizações poderosas interativas. Mas deve-se tomar cuidado. Não é porque você pode criar interatividade que isso sempre vá significar o aumento da experiência do usuário no envolvimento com a Visualização da Informação e não comprometer a essência da comunicação visual, abandonando um design estático apenas para a novidade da criação de interatividade (KIRK, 2012a).

Em alguns cenários que possuem estruturas de dados mais complexas, a capacidade de selecionar, filtrar, excluir ou modificar certas variáveis é uma forma valiosa de deixar o usuário interagir com diferentes fatias de dados. Além disso, agrupamento e classificação são opções comuns de instalações para a extração de novos insights. Um outro aspecto é a

animação. Quando temos dados baseados em séries de tempo, há um grande potencial para nós retratarmos nossa visualização através de animação, criando um cenário inconstante de dados. Na medida em que se desenrola uma história dos dados convincentes, o uso de recursos como Play, Pause, e reposição pode ser melhorado através da oferta de sliders de tempo controláveis manualmente.

Dentro desse contexto será explicado alguns exemplos de gráficos e suas aplicações na [tabela 1](#):

Tabela 1 – Exemplos de Gráficos e suas aplicações

Comparar Categorias	Avaliação de Hierarquias	Mudançãs ao longo do Tempo	Plotagem e Relacionamento	Geoespacias
Gráfico de Pontos	Gráfico de Pizza	Gráfico de Linhas	Scatter Plot	Mapa Choropleth
Gráfico de Barras	Gráfico de Barras Empilhadas	Sparklines	Bubble Plot	Mapa Gráfico de Pontos
Gráfico de Barras Flutuante	Gráfico de Pizza Quadrada	Gráfico de Área	Scatter Plot Matrix	Mapa Isarithmic
Gráfico de Barras Pixelizada	Tree Map	Gráfico de Horizonte	Heatmap	Mapa de Fluxo de Partículas
Histograma	Diagrama de Empacotamento de Círculos	Gráfico de Área Empilhado	Coordenadas Paralelas	Cartograma
Slopegrah	Hierarquia Bolhas	Gráfico de Fluxo	Rede Radial	Cartograma Dorling
Gráfico Radial	Hierarquia Árvore	Gráfico de Velas	Diagrama de Redes	Mapa de Conexão de Rede
Gráfico Glyph		Gráfico de Código de Barras		
Diagrama Sankey		Flow Map		
Gráfico de Tamanho da Área				
Múltiplos Pequenos				
Word cloud				

Fonte - Produzida pelo autor.

A melhor estratégia é certificar que a técnica utilizada para representar os seus dados está funcionando de forma eficaz e não enganar o usuário ou o leitor. Faça todas as suas escolhas de representação retratar com precisão os valores de dados eles são associados.

2.6 Combinando Técnicas de Visualização

. A combinação de técnicas exigem estratégias, uma das principais abordagens para lidar com a complexidade visual envolve informações facetadas, como por exemplo:

- justapondo vistas coordenadas lado a lado;;
- sobrepondo camadas em uma única visualização;

Nos dois casos, os dados precisam ser particionados nas visualizações ou camadas. Várias exibições justapostas lado a lado estão espalhadas no espaço, uma alternativa para uma visão em mudança onde as informações apresentadas ao usuário são espalhadas ao longo do tempo. A escolha de design multiforma para coordenar visões justapostas é usar uma codificação diferente em cada uma para mostrar os mesmos dados. O raciocínio é que nenhuma codificação visual única é ideal para todas as tarefas possíveis. Uma ferramenta multiformas pode assim suportar mais tarefas ou uma alternância mais rápida entre tarefas do que uma ferramenta que apenas mostra uma única visualização. A escolha de coordenação de pequenos múltiplos envolve a participação dos dados entre as vistas. (MUNZNER, 2014)

3 Metodologia

Neste Capítulo, é apresentada a metodologia utilizada no processo de pesquisa realizado neste estudo.

3.1 Revisão Sistemática da Literatura

A revisão sistemática da literatura é uma técnica que busca executar revisões amplas da literatura referentes a um assunto, apreciando os resultados de forma imparcial e demonstrando os critérios de seleção, com a finalidade de que o pesquisador que utilizar a revisão possa avaliar seu nível, e caso necessite, possa executá-la novamente. As primeiras técnicas formais da combinação dos resultados de diferentes estudos foram elaboradas e publicadas no *British Medical Journal*, pelo matemático britânico (PEARSON, 1904).

Durante a revisão foram utilizadas as seguintes etapas:

- Definição do tema da pesquisa: Nessa etapa foi definido o tema, seu objetivo e a extensão da pesquisa;
- Pesquisa na literatura: Nessa etapa foi efetuada busca de artigos em bases eletrônicas através de uma string de busca sobre o tema para iniciar a leitura. Nessa leitura foi iniciada uma pré-seleção dos artigos;
- Artigos não selecionados: são aqueles que objetivamente não possuem os critérios de inclusão, foram excluídos durante a fase de seleção e apenas quantificados;
- Artigos selecionados: são aqueles que possuem evidências dos critérios de inclusão, foram incluídos na fase de seleção. São registradas todas as referências completas destes artigos;
- Artigos excluídos: são aqueles que, após leitura e avaliação do texto completo, não atendem aos critérios de inclusão, são excluídos na etapa de extração; e
- Artigos incluídos: são aqueles que, após avaliação do texto completo, atendem aos critérios de inclusão, mantêm-se durante todo o processo e são utilizados na etapa de extração.

3.2 Objetivos da Revisão

A revisão executada tem como objetivo obter o conhecimento do cenário referente a utilização das técnicas de Visualização da Informação na Genética, com isso, compreender

profundamente o campo de pesquisa da Visualização da Informação aplicado à exibição de dados sobre Genética.

3.3 Questões de Pesquisa

Após diversas discussões e pesquisas prévias, chegou-se à seguinte questão principal de pesquisa:

Como as ferramentas que utilizam técnicas de Visualização da Informação aplicadas ao contexto de controle de exibição de dados sobre genética foram desenvolvidas?

Esta pergunta principal pode ser expandida nas seguintes questões:

- Qual a linguagem de programação utilizada na construção deste tipo de ferramenta?
- Quais são as técnicas de visualização que são utilizadas neste tipo de ferramenta?
- Qual o formato dos dados que são utilizados no desenvolvimento neste tipo de ferramenta?

3.4 Criação do protocolo

Nesta seção, é descrita a montagem do protocolo para direcionar a presente revisão. São detalhadas as seguintes etapas:

- definição da estratégia de busca;
- definição das bases de dados;
- definição dos critérios de inclusão e exclusão; e
- definição do processo de extração de dados com o formulário de extração e do processo de síntese dos estudos.

3.4.1 Estratégia de busca

Os termos iniciais utilizados para busca na revisão sistemática foram:

- visualization, genetics, fish;
- visualization, genetics, fish, e heritage e
- visualization e aquaculture

Utilizando-se essas palavras, não se trouxe o resultado satisfatório para realização de uma pesquisa.

Operadores lógicos podem ser utilizados nas pesquisas avançadas das ferramentas de busca, fazendo com que a pesquisa seja mais seletiva, e definindo uma instrução chamada *string* de busca. Nessa revisão foi utilizada essa *string* de busca:

(visualization) AND (genetics)

A data de início da pesquisa dos artigos ocorreu em 20/12/2020, foram considerados como estudos válidos artigos publicados na língua inglesa em periódicos e conferências do ano de 2016 a 2021. Como a área de tecnologia evolui rapidamente utiliza-se os últimos cinco anos anterior a pesquisa.

3.4.2 Bases de dados

As buscas iniciais foram realizadas utilizando *strings* de busca nas bases de dados IEEE Xplore Digital Library e ACM Digital Library.

A seleção destas buscas foi realizada considerando a relevância e a identificação com a área de pesquisa sobre as Técnicas de Visualização sobre Genética de Peixes.

3.4.3 Critérios de inclusão e exclusão

Nessa fase são definidos os critérios de avaliação de quais estudos avançam para a próxima etapa da revisão. Os critérios devem ser definidos a partir das questões de pesquisa. Outros critérios foram utilizados para inclusão como somente artigos da língua estrangeira, restrição área.

3.4.3.1 Critérios de inclusão

Na revisão foram utilizados os seguintes critérios de inclusão assim definidos:

- Coerência dos estudos com foco do tema – Foram incluídos na revisão os artigos que possuíam relação com o tema Visualização da Informação aplicada à exibição de dados sobre genética; e
- Demonstração de uma aplicação prática – Foram selecionados e identificados todos os artigos que demonstraram uma aplicação prática e completa referente ao foco do trabalho.

3.4.3.2 Critérios de exclusão

Na revisão foram utilizados os seguintes critérios de exclusão assim definidos:

- Falta de coerência dos estudos com o foco do temas – artigos retornados na busca, porém, que não apresentavam coerência com o tema da revisão, visualização da informação e genética, foram excluídos. Artigos que tratavam de somente um dos temas, também foram excluídos;
- Não é um artigo – resultados retornados que não poderiam ser considerados estudos foram desconsiderados. Dentre eles, estão artigos de opinião, índices de anais de eventos, entre outros;
- Estudo não está disponível completo no idioma Inglês – estudos retornados na pesquisa, como o resumo disponível na língua inglesa, porém, com o texto completo somente disponível em outro idioma foram excluídos da revisão pela impossibilidade de leitura; e
- Estudo não está disponível para leitura.

3.4.4 Extração de Dados

Na etapa de extração de dados, a questão de pesquisa deve ser respondida através da análise dos artigos selecionados na etapa anterior, mediante um formulário de extração previamente elaborado com questões que detalhem aspectos da pergunta de pesquisa principal. Os dados devem ser registrados de maneira que fiquem vinculados aos artigos a que se referem, Esse registro pode ser realizado utilizando ferramentas específicas ou métodos mais simplificados, como uma planilha. Ao se executar a leitura completa dos artigos, as possíveis respostas ao formulário de extração são registradas.

Para o presente estudo, as seguintes questões compuseram o formulário de extração:

- Tipos de genética: através desta pergunta, busca-se obter qual o tipo de genética tratada pelo artigo, descoberta de doenças, reprodução de animais, reprodução de vegetais, descobertas de medicamentos para cura de doenças ou estudos para outros objetivos;
- Técnicas de visualização utilizadas: identificar quais técnicas de visualização e quais recursos gráficos são utilizados no estudo e na construção das ferramentas, e quais objetivos foram alcançados com essas técnicas?;
- Fonte de dados utilizados nas visualizações: identificar quais são as fontes de dados utilizadas na construção das visualizações; e
- Linguagens utilizadas no desenvolvimento: através dessa pergunta foram analisadas as linguagens de programação utilizadas para verificar as tendências.

3.4.5 Síntese dos estudos

A síntese e sumarização das informações coletadas foram reunidas e expostas através de tabelas, de forma a mostrar a divisão dos artigos baseada nas perguntas que compõem o formulário de extração.

Sempre que possível foram utilizados gráficos e recursos visuais para exposição da informação, utilizando as boas práticas de visualização, de modo a facilitar a compreensão dos dados.

3.5 Ferramenta StArt

Para auxiliar essa revisão foi utilizada a ferramenta StArt, desenvolvida pelo Laboratório de Pesquisa em Engenharia de Software (LaPES) da Universidade Federal de São Carlos (UFSCar) (FABBRI et al., 2016).

A ferramenta possibilita aos pesquisadores preencher todo o protocolo de uma revisão sistemática de forma facilitada. Ela possui informações sobre todas as fases presentes em uma revisão, desde o objetivo, questão de pesquisa, busca, estratégias de seleção, critérios de inclusão e exclusão, formulário de extração (FABBRI et al., 2016).

3.6 Condução

No processo de revisão sistemática é comum que ocorra um aprimoramento nas strings de busca a fim de melhorar o seu resultado.

De maneira preliminar, foram obtidas as seguintes quantidades de estudos por base de dados, após a fase de identificação (Tabela 2)

Tabela 2 – Resultados — Fase de Identificação

Base de dados	Estudos identificados
IEEE Xplore Digital Library	5
ACM Digital Library	0
Total	5

Fonte: Produzida pelo autor.

Durante o processo de revisão, foram sendo incluídos operadores lógicos, e definição de quais campos do artigo seriam verificados, nos termos definidos na string.

Resultando na seguinte *string* de busca:

(visualization) AND (Genetics)

Com essa string, os resultados obtidos pelo processo estão apresentados na tabela 3:

Tabela 3 – Resultados — Fase de Identificação — Após alterações

Base de dados	Antes das alterações	Depois das alterações
IEEE Xplore Digital Library	5	147
ACM Digital Library	0	025
Total	5	172

Fonte: Produzida pelo autor.

3.7 Análise e discussão dos resultados

Neste capítulo são apresentados e discutidos os resultados obtidos ao aplicar o protocolo de revisão nas bases científicas e extrair seus dados.

3.8 Estudos

A quantidade de estudos obtida ao final da revisão, de acordo com a classificação pode ser observada na Tabela 4.

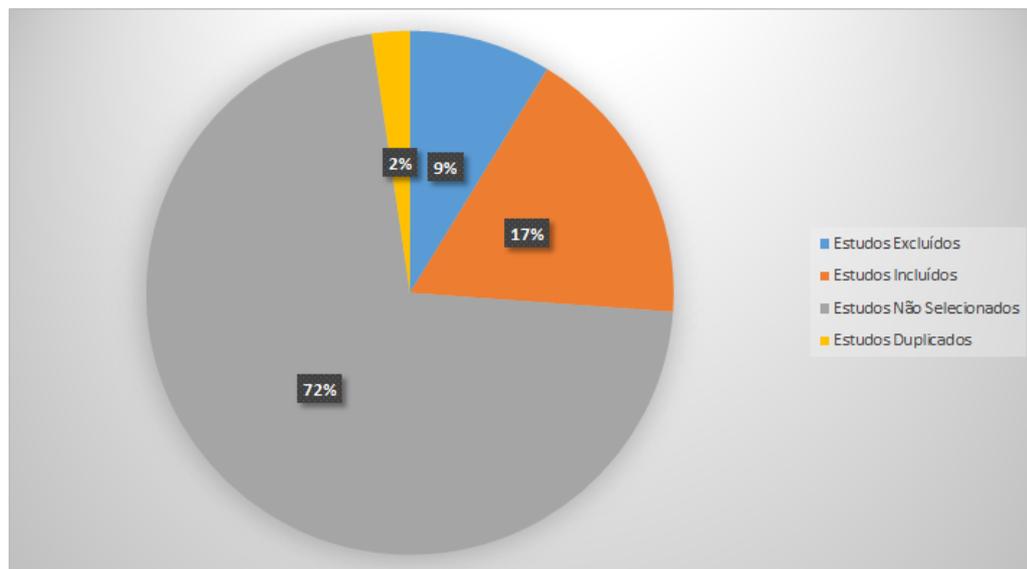
Tabela 4 – Quantidades de estudos durante as fases da revisão

Categorias	Quantidades
Estudos identificados	172
Estudos não selecionados	123
Estudos selecionados	45
Estudos excluídos	15
Estudos Duplicados	04
Estudos incluídos	30

Fonte: Produzida pelo autor.

Na Figura 1 o total de 172 estudos é dividido de acordo com a classificação final de cada trabalho, após a realização da presente revisão.

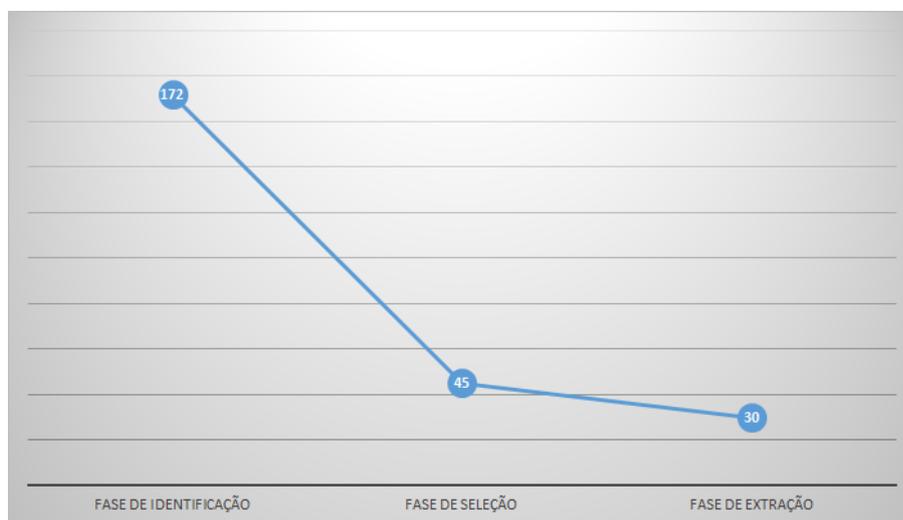
Figura 1 – Classificação final dos estudos



Fonte — Produzida pelo autor.

Na Figura 2 pode-se visualizar a evolução da quantidade de estudos aceitos em cada fase da revisão sistemática.

Figura 2 – Evolução da quantidade de estudos



Fonte — Produzida pelo autor.

3.8.1 Estudos excluídos

Um total de 15 artigos foram excluídos na fase de extração, pois, mesmo após a análise inicial do título e resumo, durante a leitura completa, foi verificado que não atendiam aos critérios de inclusão, esse artigos encontram-se no apêndice B deste trabalho.

3.8.2 Estudos Não Selecionados

Um total de 123 artigos não foram selecionados após a leitura, não foram encontrados informações sobre o foco da pesquisa.

3.8.3 Estudos incluídos

Os dados detalhados e referências dos 30 estudos analisados são apresentados no Apêndice A deste trabalho.

3.9 Tipos de Estudos sobre Genética

O assunto genética é extremamente abrangente. Dessa forma, desejou-se identificar de maneira geral qual o tipo de assunto sobre a genética é tratada em cada artigo. O conjunto maior de estudos sobre genética foi a busca pela origem de algumas doenças e medicamentos para tratamento das mesmas. Na tabela 5 foram especificado os resultados.

Tabela 5 – Ocorrência de tipos de estudo sobre genética

Tipo de Estudo sobre Genéticas	Ocorrência
AVC	1
Estudos sobre Doenças	12
Reprodução de Animais	2
Reprodução sobre Vegetais	1
Estudos sobre Medicamentos para Tratamento de Doenças	1
Estudo sobre Hereditariedade	1
Reguladores que gerencia a defesa imunológica	1
Estudo sobre Ancestralidade	2
Analise de geral de genetica de plantas, humanos, animais	1
Genética Celular	4
Estudo da presença de microrganismos em ambientes Microorganismos	1
Estudos Fatores Genéticos sobre os genes	2
Diversidade entre o mapa do Genótipo e o Fenótipo	1

Fonte: Produzida pelo autor.

3.10 Técnicas de Visualização

Foram utilizadas várias técnicas de visualização nos softwares abordados nos artigos selecionados, na tabela 6 são demonstrados a quantidade que elas ocorreram.

Tabela 6 – Técnicas visuais mais utilizadas

Técnica	Ocorrências
Barras	8
gráfico de dispersão	7
Mapas	1
Linhas	5
Linha Interligadas	1
Sunburst	2
mapa de calor	4
treemap	1
Histogramas	2
Grafo	1
Agrupamento Hierárquico	1
Arvore Multivariada	1
Bubble Tree	1
Hierarquia Radial	1
Visualização 3D	1
Diagramas de Circulos	1

Fonte: Produzida pelo autor.

Em se tratando de dados sobre genética, os quais geralmente são dados complexos as técnicas de visualização mais utilizadas são os gráficos de barras e de dispersão que têm a finalidade de comparar essas informações para um entendimento sobre tendências. Há, também, gráfico de linhas, para demonstrar uma evolução ou regresso. Nas técnicas também foram utilizadas ferramentas como o D3JS que é uma biblioteca JavaScript para produzir visualizações de dados dinâmicas e interativas, tecnologias como bootstrap, highcharts e google charts.

3.11 Objetivos da Visualização

De acordo com os objetivos da visualização utilizada e da ferramenta desenvolvida, os artigos trataram na sua maioria:

Estudos sobre os doenças genéticas: O objetivo era entender como algumas doenças que são hereditárias surgiam ocasionando aquela doença, foram encontrados 12 artigos que tratavam desse assunto;

Estudos sobre drogas para cura de doenças: O objetivo era através da genética descobrir medicamentos para tratamentos de algumas doenças, foi encontrado 1 artigo que tratava desse assunto;

Estudos sobre reprodução de animais e vegetais: O objetivo era utilizar a genética para melhorar a reprodução desses seres, que as espécies tivessem um melhor rendimento

como resultado, foram encontrados 2 artigos que tratavam desse assunto.

3.12 Formato de dados

Foram utilizados alguns formatos de dados para geração das visualizações. O mais utilizado foi o banco de dados, o qual possui recursos que facilitam essa busca dos dados. Mas alguns usaram arquivos CSV, na Tabela 7 são discriminados os formatos.

Tabela 7 – Fontes de dados utilizadas

Formato de dados	Ocorrências
Banco de dados	16
Json	2
CSV	1
Mysql	2
Mongodb	1
Não Especifica	08

Fonte: Produzida pelo autor.

3.13 Desenvolvimento e linguagem

As linguagens de programação são ferramentas utilizadas para inserir regras que serão executadas pelos computadores com objetivos definidos. Foram levantadas as linguagens de programação mencionadas nos artigos, os dados estão apresentados na Tabela 8.

Tabela 8 – Linguagens de programação

Linguagem	Utilizações
Python	7
C#	1
Javascript	3
PHP	1
Android	1
Html	1
Flash	1
C++	1
Não especificadas	14

Fonte: Produzida pelo autor.

A linguagem mais utilizada foi o Python. Em alguns artigos, foi utilizado o JavaScript juntamente com as linguagens para desenvolvimento web. Algumas bibliotecas também foram utilizadas como Dc.js e tecnologias como Bootstrap, Highcharts e Google Charts.

3.14 Relevância do estudo

As bases eletrônicas têm um papel importante na pesquisa pelo grande número de artigos disponíveis para estudo. O número de artigos selecionados, num total de 30 faz com que a revisão possua uma grande base de informação para ser aplicada na pesquisa.

3.15 Limitações do estudo

Para busca dos artigos foram utilizadas máquinas de busca das próprias bibliotecas e foram detectados em diversos artigos a falta de informações de extrema relevância para a pesquisa como qual linguagem foi utilizada, que tipo de dado e muitas vezes qual a técnica de visualização utilizada, sendo que a técnica pode ser classificada pela identificação da visualização. Mas as outras informações, não. A ausência dessas informações de dados enfraquece as tendências encontradas, mesmo com um número relevante de artigos analisados. Alguns artigos também não estavam acessíveis para leitura.

3.16 Tendências

Diversas tendências puderam ser averiguadas durante a execução da presente revisão. Dentre elas, podem ser citadas:

- Estudo Genético sobre doenças: os artigos fazem referências a várias doenças, com um destaque para o câncer;
- Técnicas de visualização: os artigos na maior parte utilizam-se de barras e dispersão;
- Objetivos da visualização: Os objetivos na utilização dessas técnicas são comparações com a utilização de barras e dispersão para que além da comparação mostre os relacionamentos;
- Linguagens de programação: As linguagens Python e JavaScript foram as mais utilizadas. Essas linguagens têm grande poder de desenvolvimento na web, e possuem aplicativos baseados no android que possibilitam o uso de mobilidade que é muito comum na coleta de dados a serem utilizados em visualizações e algumas bibliotecas baseado no Java Script que apoia o desenvolvimento; e
- Formato dos dados: : A maioria dos dados visualizados são originados de uma estrutura de banco de dados, que possui um melhor gerenciamento, mas existem dados que estão no padrão CSV, Json.

A pesquisa demonstrou que a maioria das ferramentas empregadas são para WEB, com recursos de bibliotecas como JavaScript e Frameworks e utilizando de banco de dados. Em

alguns casos, são utilizadas aplicações baseadas em Android com a finalidade de usar a mobilidade para obtenção de dados que servirão de base para a determinadas visualizações. O uso de sistemas WEB torna a aplicação mais acessível e escalável e possibilita troca de informações com outras bases de dados de interesse, que, no caso de genética, possuem grandes bancos de dados com essas informações. O uso de ferramentas visuais na área da genética é de muita importância na identificação de doenças hereditárias, na busca de medicamentos através de informações genéticas e na reprodução de animais e vegetais para um melhor desempenho nas gerações subsequentes.

3.17 Considerações Finais

O objetivo da revisão foi obter o entendimento de como representar os dados sobre genética, e que o estudo auxilie no desenvolvimento da ferramenta de visualização que será utilizada inicialmente na Unesp de Jaboticabal e poderá ser utilizada por produtores e indústrias do ramo de reprodução de peixes. Na Universidade existe um Centro de Aquicultura com laboratório de genética e há falta de ferramentas de Visualização que auxiliem os doutores e estudantes de Biologia no entendimento da melhoria genética na reprodução de peixes, estiveram envolvidos dez pesquisadores. O objetivo é encontrar através da análise dos dados a melhoria genética resultando em descendentes mais resistentes a doenças e com melhores medidas e peso. Quais os ambientes de desenvolvimento das espécies possuem uma melhor produção, que características eles possuem que podem ser determinantes. O processo hoje é feito através de mapeamento das medidas de forma manual, e, depois são gerados alguns gráficos em separado para entendimento da melhoria da reprodução. A pesquisa objetivou a investigar a área de pesquisa da Visualização da Informação aplicada à exibição de dados sobre genéticas através da execução de uma revisão sistemática de literatura, que foi realizada na bases do IEEE e ACM, e que, depois do refinamento chegou a 30 artigos de trabalho.

Para o desenvolvimento desse trabalho, a presente revisão contribuiu mostrando tendências que contribuem no processo de construção, como o uso de linguagens, e o tipo de dado usado para ser visualizado. Tendências que contribuem com a definição do objetivo da ferramenta, como o foco no estado geral na exibição de dados de genética.

4 Ferramenta de Visualização Genética de Peixes

Com base nas pesquisas realizadas foi desenvolvido uma ferramenta de visualização para análise dos dados que será descrita nesse capítulo.

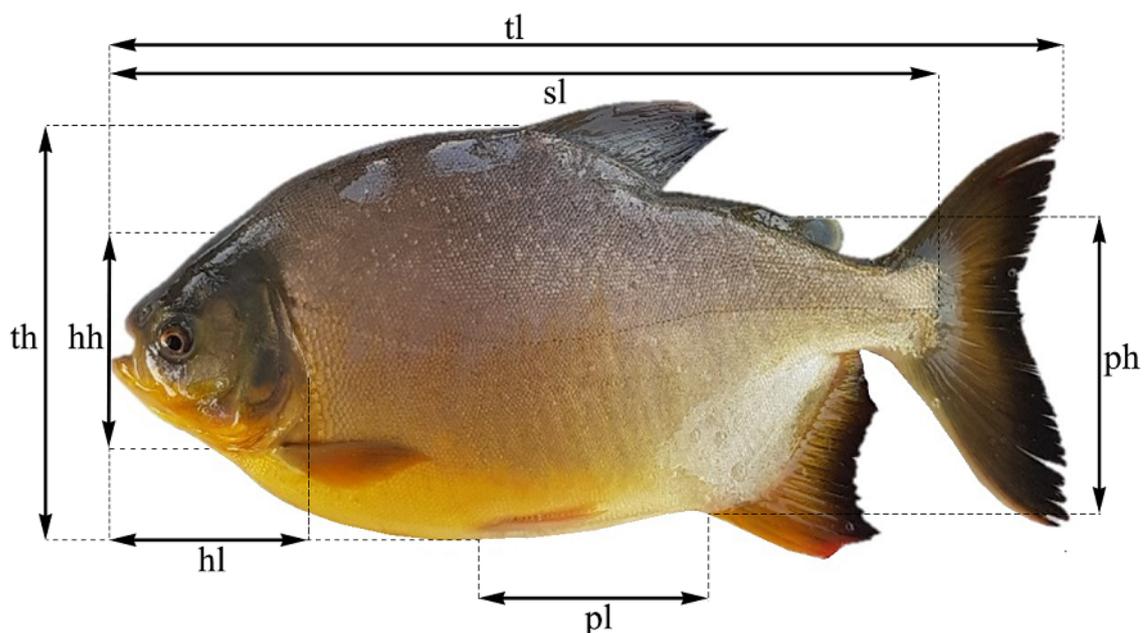
4.1 Cenário Atual

As pesquisas com a extração de características dos peixes na Unesp em Jaboticabal são feitas através de fotos de celulares. Geralmente os animais são colocados em tampas de isopor no chão, e utilizada a mesma medida de distância para que as fotos possuam o mesmo padrão. As medidas que são retiradas dos peixes acontecem de forma manual, o que leva um tempo considerável para a execução da tarefa. Através dos dados que estão em planilha de Excel são gerados alguns tipos de gráficos para análise dos pesquisadores. A Figura 3 é uma foto de uma espécie de peixe chamada Pacu cujo o nome científico é "Piaractus mesopotamicus", as informações contidas na imagem são tamanho total (tl), comprimento padrão (sl), tamanho da cabeça (hl), comprimento da pelve (pl), altura total (th), altura da cabeça (hh), altura na região posterior (ph), além do Pacu, o Tambaqui e a Tilápia são espécies que são analisadas.

Nesse cenário foi desenvolvida uma ferramenta que possibilite :

- Utilizar os dados extraídos nas planilhas com as medidas e outras informações;
- Utilizar um conjunto de técnicas de visualização para auxiliar os pesquisadores na identificação dos melhores conjuntos de espécies para retirada de filés;

Figura 3 – Foto do Pacu com as referências para medidas



Fonte — Produzida pelos pesquisadores da Unesp de Jaboticabal.

Depois que os dados foram extraídos são transportados para uma planilha que está na Figura 4. Essas informações são utilizadas para descoberta das melhores espécies.

Figura 4 – Imagem da Planilha de Fenótipos

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	tag	id	dam	sire	caixa	tank	ambcom	amb	Nascimento	Envio	Manejo	idade total	idade pós envio	iw	fw1	fw2	fw	hh	hl	tl	sl	th	ph	pl	adg	EBV	Sexo
2	942822	39	1	17	11	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	27,3	0	995	995	76,8	72,4	299,5	226,2	136,2	101,3	75	4,3	-4.666.504.981	M
3	942766	40	1	17	11	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	48,3	0	1050	1050	77,4	70,5	290,4	222,3	134,8	98,6	70,6	3,8	-5.564.135.621	X
4	942451	41	1	17	11	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	31,5	0	1220	1220	84,8	68,8	316,6	248,2	148,1	105,9	83,3	4,7	358.096.635	
5	942436	42	1	17	15	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	58,3	0	1030	1030	84,8	72,7	310,4	239,1	141,4	106	74,4	3,6	-8.133.582.650	
6	942270	43	1	17	15	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	45,5	0	1150	1150	77,4	81	310,9	234,5	143,4	105,4	72,8	4,3	-4.337.499.901	F
7	942131	44	1	17	15	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	48,5	0	1150	1150	81,6	74,9	314,7	247,7	147,9	107,2	74,8	4	-2.989.910.065	
8	942322	45	1	17	16	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	43,3	0	1180	1180	75,1	70,1	303,3	235	141	101,4	79,9	4,4	-3.102.735.965	F
9	942138	46	1	17	16	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	38,1	0	1240	1240	90,3	80	324,7	248,8	151,8	105,3	78,9	4,3	1.492.001.752	
10	941484	47	1	17	16	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	33,5	0	885	885	72,4	71	270,5	201,9	136,3	95,2	71,7	3,9	-8.811.952.635	
11	942823	48	1	17	16	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	23,1	0	960	960	80,6	76,7	285	210,7	136,5	93,7	67,2	4,1	-2.760.805.612	
12	942469	49	1	17	16	4	2	2	17/11/2017	26/10/2018	01/04/2019	500	157	41,6	0	1100	1100	78,1	71,4	309	240,9	143,3	106,3	81,1	4,2	-4.846.968.281	
13	942472	50	1	17	16	5	2	2	17/11/2017	26/10/2018	10/04/2019	509	166	41,9	0	890	890	79,2	67,5	298	232,9	136,4	99,5	74,2	3,6	-8.438.676.168	
14	942628	51	1	18	11	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	44,5	0	1170	1170	80,3	74,8	306,5	235,4	152,9	104,8	77,9	4,3	-4.317.886.258	
15	620170	52	1	18	11	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	50,9	0	880	880	76,1	71,5	285,7	214,4	134,1	100,1	64,2	3,4	-11.250.030.958	
16	942877	53	1	18	11	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	61,3	0	1140	1140	82,3	71,9	309,3	239,3	141,5	104,2	69	4	-8.540.411.565	M
17	941454	54	1	18	11	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	32,2	0	790	790	76	68,4	267	203,6	121	90,9	61,8	3,6	-9.776.871.102	M
18	942344	55	1	18	15	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	32,4	0	1125	1125	81,6	71,9	303	234,1	146,7	103,5	75,8	4,5	-3.018.542.260	
19	942861	56	1	18	15	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	32,8	0	850	850	77	73,7	286,9	219,1	129,9	93,8	70,7	3,7	-8.330.563.340	
20	942188	57	1	18	15	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	49,8	0	1160	1160	79	74,7	298,6	227,1	145,7	103,7	73,8	4,2	-5.664.002.103	
21	942848	58	1	18	16	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	32,3	0	1180	1180	84,9	72,9	317,2	248,7	153,7	108,8	81,3	4,3	401.765.795	
22	942492	59	1	18	16	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	28,9	0	1035	1035	75,5	68,5	280,5	213,3	131	96,4	72	4,4	-4.655.934.826	
23	942601	60	1	18	16	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	44,6	0	1020	1020	76,9	69,5	299,4	230,3	142,5	104,2	68,4	3,8	-6.299.572.462	
24	942555	61	1	18	16	5	4	2	17/11/2017	26/10/2018	10/04/2019	509	166	39,3	0	1000	1000	82,7	71	297,8	224,3	135,5	98,8	71,1	3,9	-5.738.002.828	M POUCO DESENV
25	942566	62	1	18	16	4	4	2	17/11/2017	26/10/2018	01/04/2019	500	157	22,1	0	1100	1100	81,7	72,2	302,2	233,5	150	105	74,3	4,7	-1.564.521.827	
26	1046592	63	1	19	3	1	5	1	17/11/2017	01/10/2018	15/04/2019	514	196	30,5	660	0	660	74,8	68,2	272,3	200,8	124,6	89,1	60,3	4,4	-1.554.414.787	
27	1045934	64	1	19	3	4	5	2	17/11/2017	26/10/2018	01/04/2019	500	157	39	0	1255	1255	82	74,6	307,1	232,6	151,3	108,6	81,6	4,6	1.634.815.725	F
28	1045876	65	1	19	3	2	5	1	17/11/2017	01/10/2018	03/04/2019	502	184	16,8	480	0	480	66,8	60,4	235,3	172,3	112,4	78,5	57,1	4,6	-2.435.283.641	
29	1045891	66	1	19	3	5	5	2	17/11/2017	26/10/2018	10/04/2019	509	166	38,6	0	990	990	83	75,9	295,7	224,1	140,5	99,4	73,1	3,9	-3.251.976.810	M
30	1045945	67	1	19	4	3	5	1	17/11/2017	01/10/2018	04/04/2019	503	185	50,9	510	0	510	71,2	62,4	261,4	200,5	112,4	83,2	58,6	4	-4.182.131.505	
31	1046613	68	1	19	4	1	5	1	17/11/2017	01/10/2018	15/04/2019	514	196	39,8	720	0	720	73,4	68,1	276,8	210	127,7	89,6	63,1	4,4	-4.247.382.619	
32	1046566	69	1	19	4	5	5	2	17/11/2017	26/10/2018	10/04/2019	509	166	34,9	0	1045	1045	80,7	71	286,1	218,7	128,5	97,5	61,7	4	-1.056.468.712	M
33	1045890	70	1	19	4	2	5	1	17/11/2017	01/10/2018	03/04/2019	502	184	32,2	830	0	830	76,7	70	270	207,9	132,6	94,8	65,3	5,4	1.658.994.069	

Fonte — Produzida pelos pesquisadores da Unesp de Jaboticabal.

4.2 Tecnologias Envolvidas

Nesta seção estão apresentadas as tecnologias escolhidas que foram utilizadas para o desenvolvimento da ferramenta.

4.2.1 CSharp

O CSharp é uma linguagem de programação moderna, orientada a objeto e de tipo seguro. Ela permite que os desenvolvedores criem muitos tipos de aplicativos seguros e robustos que são executados no ecossistema do .NET. O CSharp tem suas raízes na família de linguagens C e os programadores em C, C++, Java e JavaScript a reconhecerão imediatamente. É uma linguagem de programação orientada a objeto, fornece construções de linguagem para dar suporte direto a esses conceitos, tornando a uma linguagem natural para criar e usar componentes de software. Desde sua origem, o CSharp adicionou recursos para dar suporte a novas cargas de trabalho e práticas de design de software emergentes (MICROSOFT, 2021).

4.2.2 Arquitetura .Net

Os programas em CSharp são executados no .NET, um sistema de execução virtual chamado Common Language Runtime (CLR) e um conjunto de bibliotecas de classes. O CLR é a implementação da Microsoft da CLI (Common Language Infrastructure), um padrão internacional. A CLI é a base para a criação de ambientes de execução e desenvolvimento nos quais as linguagens e bibliotecas funcionam em conjunto diretamente. O código-fonte escrito em CSharp é compilado em uma IL (linguagem intermediária) que está de acordo com a especificação da CLI. O código de IL e os recursos, como bitmaps e cadeias de caracteres, são armazenados em um assembly, normalmente com uma extensão de dll. Um assembly contém um manifesto que fornece informações sobre os tipos, a versão e a cultura do assembly (MICROSOFT, 2021).

4.2.3 Java Script

JavaScript é uma linguagem leve, interpretada e baseada em objetos com funções de primeira classe, mais conhecida como a linguagem de script para páginas Web, mas usada também em vários outros ambientes sem browser, tais como node.js, Apache CouchDB e Adobe Acrobat. O JavaScript é uma linguagem baseada em protótipos, multi-paradigma e dinâmica, suportando estilos de orientação a objetos, imperativos e declarativos (como por exemplo a programação funcional). O padrão JavaScript é ECMAScript. Desde 2012, todos os navegadores modernos possuem suporte total ao ECMAScript 5.1. Navegadores mais antigos suportam pelo menos ECMAScript 3. Em 17 de Junho de 2015, a ECMA International publicou a sexta versão do ECMAScript,

que é oficialmente chamado de ECMAScript 2015, e foi inicialmente conhecido como ECMAScript 6 ou ES6 ([JAVASCRIPT, 2021](#)). O JavaScript foi criado na década de 90 por Brendan Eich a serviço da Netscape. Essa década foi um período de revolução, pois os browsers ainda eram estáticos. O navegador mais popular dessa época era o Mosaic, da NCSA ([RAUSCHMAYER, 2014](#))

4.2.4 SQLite

SQLite é uma biblioteca que implementa um mecanismo de banco de dados SQL transacional independente, sem servidor e sem configuração. O código para SQLite é de domínio público e, portanto, seu uso é gratuito para qualquer finalidade, comercial ou privada. SQLite é um mecanismo de banco de dados SQL incorporado, a sigla SQL significa Standard Query Language, literalmente a linguagem padrão para realizar queries. Ao contrário da maioria dos outros bancos de dados SQL, a biblioteca não possui um processo de servidor separado. Ela lê e grava diretamente em arquivos de disco comuns. Um banco de dados SQL completo com várias tabelas, índices, gatilhos e visualizações está contido em um único arquivo de disco. O formato do arquivo de banco de dados é multiplataforma ([SQLITE, 2022](#)).

4.2.5 D3

O D3.js foi criado para preencher uma necessidade premente de visualização de dados sofisticada e acessível pela Web, fornece aos desenvolvedores a capacidade de criar rico conteúdo interativo e animado com base em dados e vincular esse conteúdo aos elementos existentes da página da web, possuem ferramentas para criar painéis de dados de alto desempenho e visualização sofisticada de dados, além de atualizar dinamicamente o conteúdo tradicional da Web. Embora a capacidade de criar gráficos ricos e variados seja um dos pontos fortes do D3, o mais importante para o desenvolvimento moderno da web é a capacidade de incorporar o alto nível de interatividade que os usuários esperam. O D3.js oferece aos desenvolvedores a capacidade de criar não apenas aplicativos altamente interativos, mas também aplicativos estilizados e servidos como conteúdo da Web tradicional ([MEEKS, 2018](#)). Nesse Capítulo foram detalhadas as especificações do projeto utilizadas para o desenvolvimento da ferramenta de visualização. O detalhamento passou pelo cenário atual, o processo de obtenção de dados e tecnologias que foram utilizadas.

5 Desenvolvimento da Ferramenta

Nesse capítulo será descrito sobre o funcionamento da ferramenta de visualização, ela já passou por uma análise de testes dos pesquisadores da área de biologia, aquicultura, zootecnista e veterinária.

5.1 Módulos

O desenvolvimento foi executado de forma modular, definição do tratamento dos dados, depois como seria a manutenção dessas informações e por último a geração dos gráficos. A execução das tarefas seguiu essa sequência :

- Importador: Importação dos dados que estavam em planilhas e armazenar em um banco de dados para melhor acesso e pesquisas;
- Lista de Peixes: Manutenção desse dados através de listas com possibilidade de alterações, inclusões e exclusões quando necessário;
- Dashboard: Geração dos gráficos utilizando técnicas de visualização

5.2 Importador

O recurso da funcionalidade da importação dos dados é feita através de um arquivo no padrão CSV, esses dados alimentam um banco de dados Sqlite. A formatação dos dados obedece uma estrutura fixa que foi definida pelo pesquisadores na utilização dos dados através da planilha. Na Figura 5 está apresentada a funcionalidade do importador da ferramenta.

Figura 5 – Figura da Ferramenta de Análise da Funcionalidade de Importação



Fonte — Produzida pelo próprio Autor

5.3 Lista de Peixes

O recurso da lista de peixes foi desenvolvida com o objetivo do pesquisador visualizar os dados após a importação, nessa opção há a possibilidade para fazer manutenção nos dados incluindo, alterando e excluindo. Para utilização da técnica de visualização das coordenadas paralelas, os dados chamados de EBV passaram por um processo chamado de padronização de dados. Essa técnica de dimensionamento foi utilizada para que os valores sejam centralizados em torno da média com o desvio padrão da unidade, visto que o dado do EBV é um número absoluto, o que impossibilita a plotagem dessa informação nessa técnica. A equação (5.1) demonstra o cálculo que é executado para a padronização dos dados EBV, é um cálculo matemático que redimensiona os dados para deixar na mesma escala. Na Figura 6 está apresentada a funcionalidade da lista de peixes.

$$z = \frac{x - \mu}{\sigma} \quad (5.1)$$

Figura 6 – Figura da Lista de Peixes

The screenshot shows a web browser window with the URL 'localhost/Peixes/Index'. The page title is 'Lista de Peixes'. There is a search bar with 'Morfometria' selected. Below the search bar, there is a table with the following columns: Ações, Tag, Id, Tipo, Dam, Sire, Caixa, tank, Ambcom, Amb, Nascimento, Manejo, Idade Total, IW, FW, HH, HL, TL, SL, TH, PH, PL, Adg, EBW FW, EBW FW Padronizado, EBV HH, and EBV HH Padronizado. The table contains 10 rows of data. At the bottom, there is a pagination control showing 'Mostrando 1 para 10 de 1.029 registros' and a page number '1'.

Ações	Tag	Id	Tipo	Dam	Sire	Caixa	tank	Ambcom	Amb	Nascimento	Manejo	Idade Total	IW	FW	HH	HL	TL	SL	TH	PH	PL	Adg	EBW FW	EBW FW Padronizado	EBV HH	EBV HH Padronizado
🗑️	942822	37	Pacu	1	17	21	5	70	2	12/12/2017	10/04/2019	484	22.5	810	75.2	69.4	290.2	224.3	133.7	70.7	97.1	3.8	5512566819	0.97209354307833	-0.06468853	0.00561471328995
🗑️	942503	714	Pacu	10	32	9	3	56	1	12/12/2017	04/04/2019	478	36.1	670	75.1	63.4	251.7	188.5	121.5	59.3	88.4	5	1842501911	0.370071850386144	183279775	1.5471305339027
🗑️	942769	715	Pacu	10	32	18	5	68	2	12/12/2017	10/04/2019	484	38.4	1230	85.8	78	316.9	238.1	142.1	82.8	105	4.3	12112592057	2.05473335060075	268702941	2.26560130707809
🗑️	942455	716	Pacu	10	32	9	1	56	1	12/12/2017	15/04/2019	489	24.7	710	74.2	68.8	264.8	200.6	127.9	63	92.2	4.8	4984418756	0.885458402052274	162545210	1.37273780533767
🗑️	942390	717	Pacu	10	32	11	4	24	2	17/11/2017	01/04/2019	500	43	1045	76.2	75.6	281.8	207.4	130.3	58.2	93.8	4.1	-4711267638	-0.704980372478653	0	0.00561471383402
🗑️	123590	718	Pacu	11	33	16	5	2	2	17/11/2017	10/04/2019	509	38.1	1240	90.3	80	324.7	248.8	151.8	78.9	105.3	4.3	1492001752	0.312577308927189	-0.32841655	0.00561471107180
🗑️	563898	719	Pacu	11	33	4	1	5	1	17/11/2017	15/04/2019	514	39.8	720	73.4	68.1	276.8	210	127.7	63.1	89.6	4.4	-4247382619	-0.628806665322636	-146830048	-1.2293324688444
🗑️	942054	720	Pacu	11	33	16	5	21	2	17/11/2017	10/04/2019	509	18.9	870	77.7	65.5	277.8	216.4	137.6	70.7	99.7	4	-8620282428	-1.34619837895257	-166711103	-1.396465588441
🗑️	563859	721	Pacu	11	33	4	3	11	1	17/11/2017	04/04/2019	503	41.5	540	70.3	69.7	259.9	188.1	117.7	60.1	85.1	4.3	-8031235935	-1.38080228576185	-125546062	-1.0503187117498
🗑️	563866	722	Pacu	11	33	4	3	11	1	17/11/2017	04/04/2019	503	45.1	700	76.7	71	277.7	206.8	124.6	62.1	87.6	4.9	-6788390365	-1.04570266796508	-0.80663562	0.00561470704963

Fonte — Produzida pelo próprio Autor

5.4 Dashboard

O recurso do dashboard foi desenvolvido com duas técnicas de visualização proporcionando ao pesquisador uma análise com enfoques distintos. O objetivo é que ele possa visualizar os dados genéticos de uma forma mais rápida, sugestiva e conclusiva. Para isso foi utilizado a técnica de coordenadas paralelas e diagrama de cordas, a geração dos dados através das técnicas são executadas através de filtros que a ferramenta possui para analisar diferentes cenários. Os filtros existentes na ferramenta são :

Tipo de Pesquisa - Define o tipo de pesquisa a ser analisada

- Mofometria: Analisa as características referente as medidas dos peixes;
- Resistência: Analisa o quanto os peixes possui resistência a doenças.

Espécie: Define a espécie ser analisada

- Pacu;
- Tambaqui e;
- Tilápia.

Casais: Carrega os casais por pesquisa de acordo com a planilha original que foi importada para múltiplas seleção.

Tipo de Dados: Tem comportamentos diferentes dependendo do filtro da pesquisa, na morfometria esta disponibilizado todas as medidas dos peixes conforme a Figura 3.

- IW;
- FW;
- HH;
- HL;
- TL;
- SL;
- TH;
- PH e;
- PL;

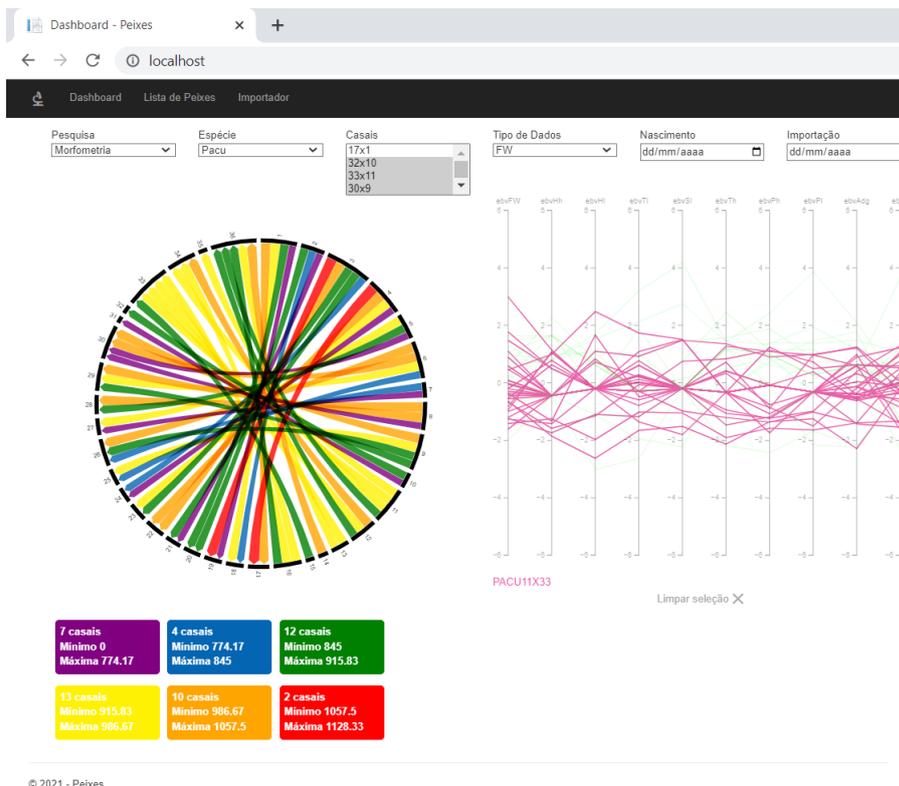
E na resistência somente essas:

- FW;
- IW.

As datas de nascimento e importação são utilizadas como filtro.

Na Figura 7 está apresentada a funcionalidade do dashboard.

Figura 7 – Dashboard



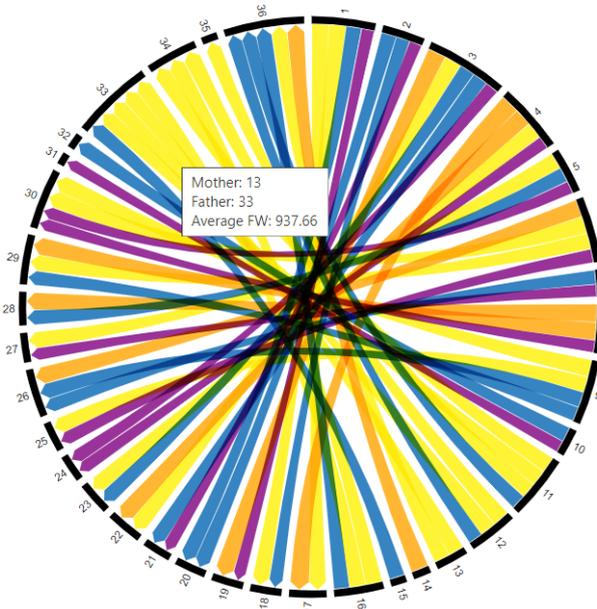
Fonte — Produzida pelo próprio Autor

5.5 Diagrama de Cordas

Essa técnica utiliza uma rede radial ou de diagramas de cordas, cujo quadro para comparar complexas relações entre valores categóricos é criado. O uso de ofertas radiais de layout dá a oportunidade de ir além das restrições de emparelhamento de eixos x e y. A chave da propriedade explicativa são as ligações que existem entre os componentes, por vezes, tamanho (espessura) e variação de cores para incorporar camadas extras de detalhes (KIRK, 2012b). Na Figura 8 está apresentada a Técnica de Visualização do Diagrama de Cordas.

A utilização dessa técnica possibilita analisar em cada fita da técnica qual é o relacionamento entre os peixes, podendo ser analisado em cada característica a média das medidas. As legendas que são classificadas e organizadas abaixo pelas medidas demonstrando a quantidade dos casais das menores medidas para as maiores auxiliando o pesquisador.

Figura 8 – Figura da Técnica de Visualização de Diagrama de Cordas



Fonte — Produzida pelo próprio Autor

5.6 Coordenadas Paralelas

A definição da técnica consiste em conjuntos paralelos que oferecem uma maneira única de explorar e analisar visualmente os conjuntos de dados. A técnica consiste em traçar todos os dados para uma série de eixos, cada um desses representam as variáveis de interesse que serão analisadas na técnica. Isto cria caminhos que mostram as conexões entre a desagregação dos valores contidos nos dados para cada variável. Eles são úteis para aprender sobre as potenciais correlações e consistências que existem nos bancos de dados. (KIRK, 2012b). A técnica Coordenadas Paralelas é uma das técnicas de visualização mais conhecidas e estudadas. Ela ainda é foco de estudo de muitos pesquisadores e de diversas variações para melhorar a análise de dados multivariada (INSELBERG, 1985). Na Figura 9 está apresentada a Técnica de Visualização das Coordenadas Paralelas.

A utilização dessa técnica possibilita analisar a variação dos índices genéticos(EBV) de cada característica podendo comparar um casal de peixes com o outro através da utilização do filtro, possibilitando a seleção dos casais, existe uma outra funcionalidade da ferramenta que é a seleção das linhas que cruzam os eixos, as quais demonstram a variação desses dados.

Figura 9 – Figura da Técnica de Visualização de Coordenadas Paralelas

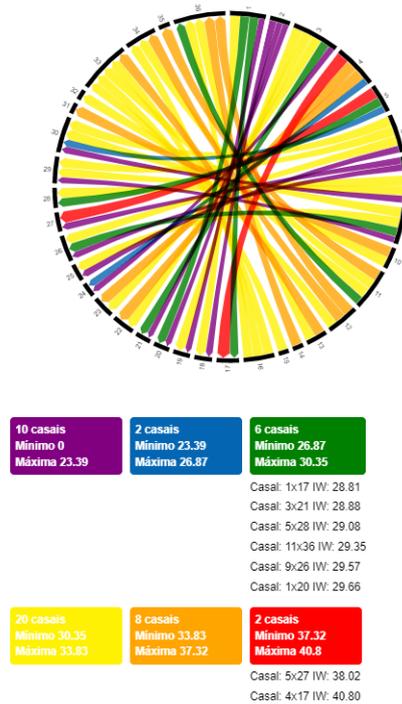


Fonte — Produzida pelo próprio Autor

5.7 Utilização da Ferramenta

Essa seção será descrita de como utilizar a ferramenta para efetuar análises dos casais de peixes. A Técnica de Visualização do gráfico de cordas utiliza os filtros de pesquisa, espécie e o tipo de dados que são as características de morfometria dos peixes. Com essa técnica é possível selecionar os melhores casais em cada característica e com a utilização da legenda que é gerada na classificação facilita a procura, essa funcionalidade quando pressionada demonstra os casais que estão inseridos nessa medida fazendo com que o pesquisador possa encontrar os casais com as melhores características com o valor máximo e mínimo das medidas. Na Figura 10 está apresentada a legenda com medidas classificadas.

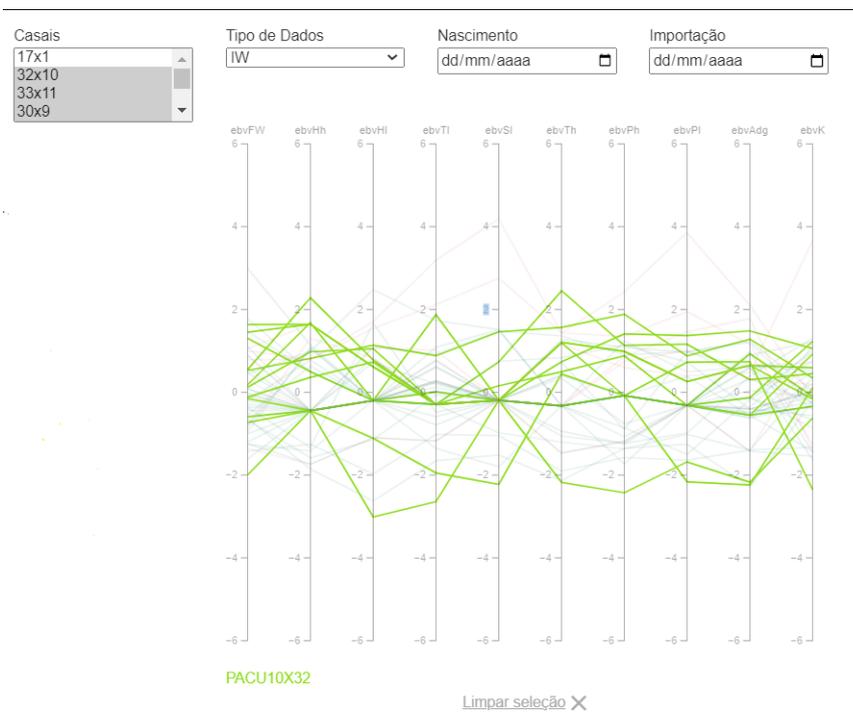
Figura 10 – Figura da Técnica de Visualização de Diagrama de Acordes com a Legenda



Fonte — Produzida pelo próprio Autor

Na Técnica de Visualização das Coordenadas Paralelas utiliza o filtro de casais podendo selecionar um ou vários, nessa visualização é possível selecionar os melhores casais através dos índices genéticos que são o EBV, existe a possibilidade do pesquisador selecionar uma ou mais linhas das Coordenadas para comparativo entre os casais, com isso a análise visual é imediata. Na Figura 11 esta demonstrada essa seleção.

Figura 11 – Figura da Técnica de Visualização de Coordenadas Paralelas com a seleção de Linhas



Fonte — Produzida pelo próprio Autor

E para concluir essa abordagem com as duas técnicas pode se comparar através das características de morfometria e os índices genéticos a existência de uma correlação entre essas informações possibilitando encontrar padrões para a escolha dos melhores casais de peixes da mesma espécie..

6 Avaliação

Para validação da ferramenta desenvolvida, realizou-se uma avaliação de usabilidade com a participação de pesquisadores das áreas de biologia, como aquicultura, zootecnia, veterinária e genética. A ferramenta inicialmente foi instalada no laboratório da Unesp de Jaboticabal e alguns pesquisadores tiveram contato com ela, mas para ser testada e analisada em virtude da pandemia foi disponibilizada no ambiente da web porque os pesquisadores estava em diferentes lugares do país e alguns deles estavam fora do país. Nesse capítulo será detalhado o processo e o questionário.

6.1 Processos

A ferramenta como dito anteriormente foi instalada no computador do laboratório de aquicultura da Unesp de Jaboticabal estado de São Paulo, os pesquisadores que contabilizaram dez pessoas começaram a ter contato com a ferramenta, alguns ajustes foram sendo efetuados nesse percurso para melhoraria da interface e dos filtros. Para que todos os pesquisadores tivessem acesso a ferramenta foi contratado um serviço de hospedagem de sites, pois eles estavam descentralizados em virtude da pandemia, havia pessoas em cidades diferentes e no exterior. Eles efetuaram testes na ferramenta para fazer análise em dois conjunto de dados diferentes, um na pesquisa de morfometria e outro na resistência. Depois efetuaram as respostas do questionário que será detalhado a seguir.

6.2 Questionário

O questionário foi preparado buscando as melhores questões para o contexto desejado. Foi utilizado questões de múltipla escolha e aberta. Utilizou se o Google Forms para facilitar a aplicação e contabilização. As questões foram distribuidas para validar a interface, usabilidade e resultado da análise. O questionário foi enviado ao coordenador do curso de pós graduação da Unesp de Jaboticabal da área de Aquicultura, ele repassou aos pesquisadores totalizando 10 pessoas que estavam em diferente cidades e no exterior. Os testes solicitados foram realizados e o questionário respondido, será demonstrado as questões e os gráficos com os resultados. Em algumas questões para entendimento e aprimoramento da ferramenta foi deixado um espaço para que os pesquisadores colocassem comentários.

6.3 Video

Para auxiliar no processo de entendimento da ferramenta foi gravado um vídeo demonstrando como utilizar as funcionalidades para que os pesquisadores conseguissem testar e efetuar uma análise dos dados que estavam no banco de dados da ferramenta.

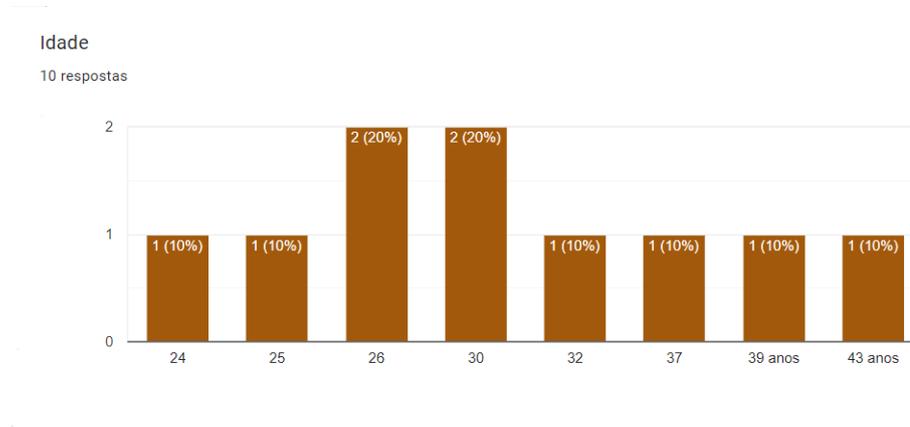
6.4 Questionário e Respostas

Nas seções 6.5, 6.6, 6.7 e 6.8 serão detalhados as questões e respostas obtidas. As perguntas foram compostas da identificação dos participantes, assuntos relacionados a interface, usabilidade e análise dos dados. Para otimizar a aplicação foi utilizado a plataforma do Google Forms.

6.5 Identificação

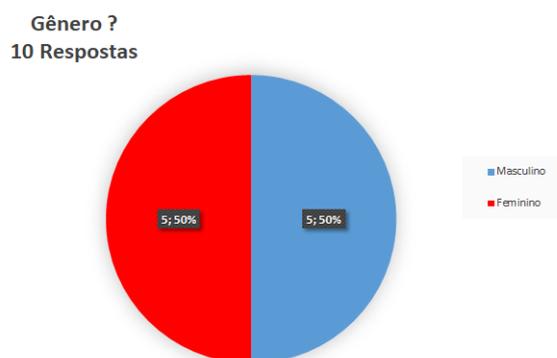
Na identificação dos avaliadores foram solicitados o nome, idade, gênero, formação, no caso da formação se esta em andamento ou concluída e email.

Figura 12 – Resultado sobre a Idade dos Pesquisadores



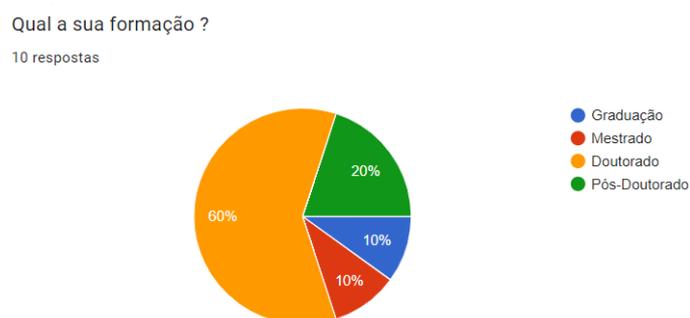
Fonte — Produzida pelo próprio Autor

Figura 13 – Resultado Sobre o Gênero dos Pesquisadores



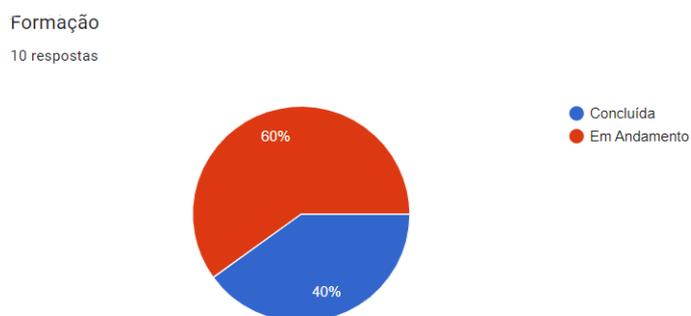
Fonte — Produzida pelo próprio Autor

Figura 14 – Resultado Sobre o Nível de formação dos Pesquisadores



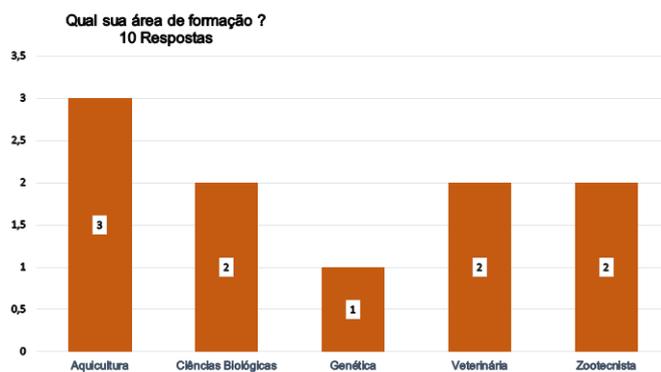
Fonte — Produzida pelo próprio Autor

Figura 15 – Resultado sobre o Estado Atual da Formação dos Pesquisadores



Fonte — Produzida pelo próprio Autor

Figura 16 – Resultado sobre a Área de Formação dos Pesquisadores

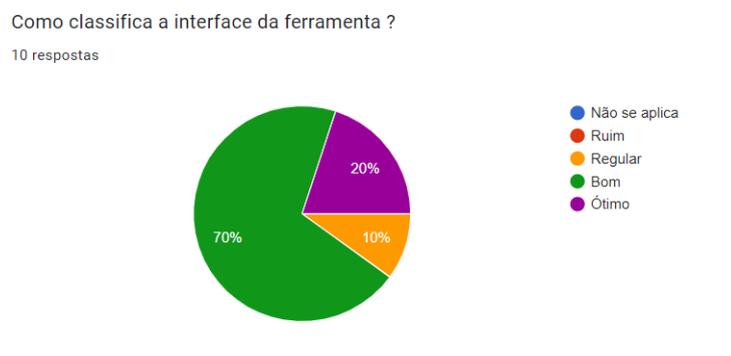


Fonte — Produzida pelo próprio Autor

6.6 Interface

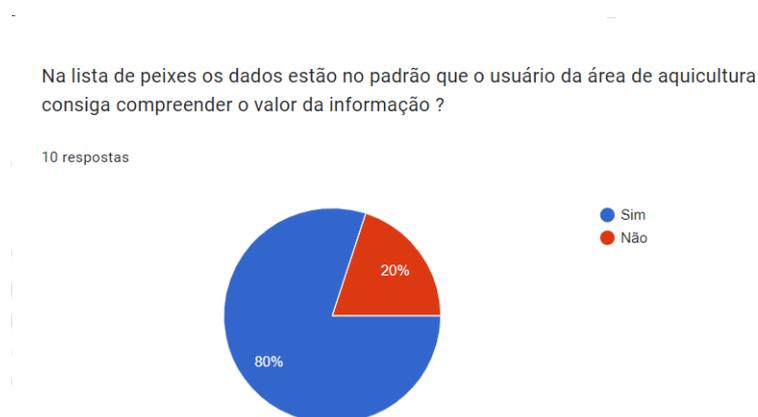
No que se refere a interface foi questionado de uma forma geral como ela se classifica conforme demonstrado na Figura 7.

Figura 17 – Resultado sobre a Interface



Fonte — Produzida pelo próprio Autor

Figura 18 – Resultado sobre a Interface da Lista de Peixes

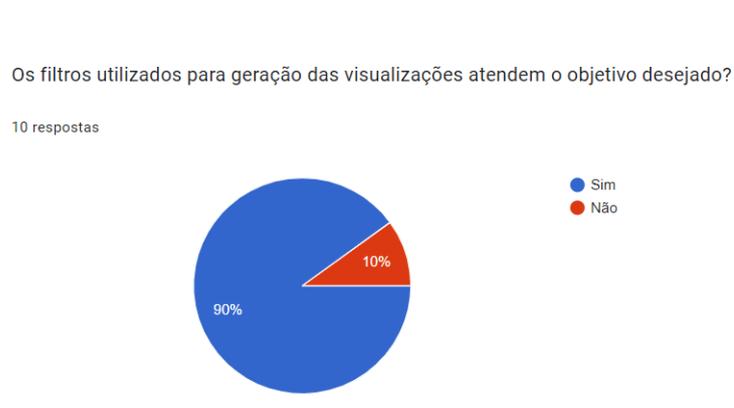


Fonte — Produzida pelo próprio Autor

6.7 Usabilidade

Referente a usabilidade foi questionado sobre os filtros que a ferramenta possui.

Figura 19 – Resultado da Usabilidade do Filtros



Fonte — Produzida pelo próprio Autor

Comentários que os pesquisadores deixaram como contribuição:

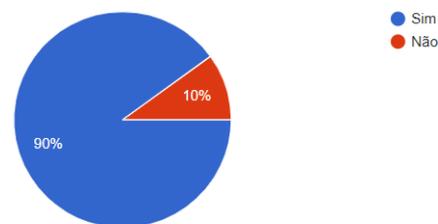
- "Na pesquisa para resistência como para morfometria não faltou nenhuma das características necessárias atualmente para seleção.";
- "A caixa de seleção casal do dashboard deveria ser substituída por uma caixa de seleção com o id do indivíduo progênie";
- "Exista alguma opção de filtro na lista de peixes";

- "Exista um campo de busca direta (como a opção localizar do Excel) nas caixas de seleção do dashboard e na lista de peixes";
- "Exista um campo de busca direta da página na lista de peixes";
- "Os filtros utilizados atendem a visualização, no entanto poderia ter cores mais distintas";
- "Acredito que a utilização dos filtros atendem o objetivo para direcionar acasalamentos com base nos melhores casais e EBVs";
- "Somente trocaria o termo Casais por Família e Pesquisa por “Característica”;
- "A ordem dos casais está invertida. Por exemplo, no filtro está casal 27x5, mas na legenda das cores que indicam mínimo e máximo, está 5x27. Pode ficar confuso para procurar, assim como a ordem dos casais";
- "Acho que as informações poderiam estar distribuídas de forma diferente, como as caixas que apresentam as informações da quantidade mínima e máxima de casais poderiam apresentar uma apresentação mais destacada.";

Figura 20 – Resultado da Usabilidade da Ferramenta

Caso utilize alguma ferramenta para análise você utilizaria essa como apoio ou substituiria ?

10 respostas



Fonte — Produzida pelo próprio Autor

Comentários que os pesquisadores deixaram como contribuição:

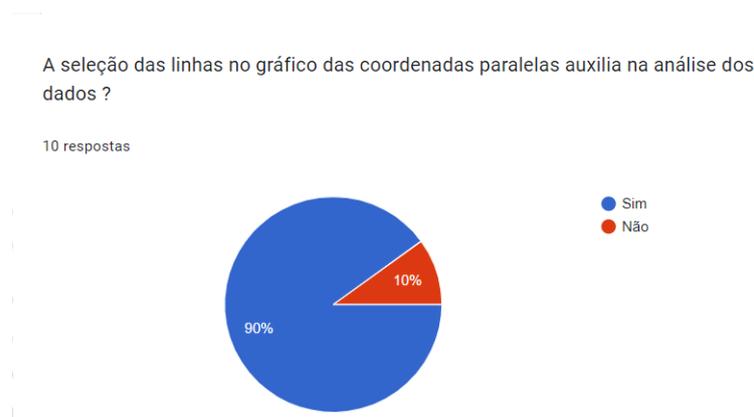
- "Apenas como apoio";
- "A facilidade para gerar os gráficos e entregar os melhores acasalamentos, sem a necessidade de estar empregando formulas ou reajustando os dados para plotagem, torna o software uma ferramenta muito amigável para trabalhar.";

- "Utilizaria após os ajustes";
- "Utilizaria como apoio, mas não substituiria e";
- "Com adequações utilizaria, mas por exemplo colocar legenda para as características de interesse. IW, FW ";

6.8 Análise

No caso da análise que é objetivo principal da ferramenta foram efetuadas várias abordagens.

Figura 21 – Resultado da Análise da Seleção das Linhas da Coordenadas Paralelas



Fonte — Produzida pelo próprio Autor

Comentários que os pesquisadores deixaram como contribuição:

- "Sim, pois seleciono o acasalamento que desejo e já busco na legenda para avaliar o desempenho";
- "Sugiro incorporar mais informações (ex: id do indivíduo) ao passar o cursor sob as linhas do gráfico de coordenadas paralelas";
- "A possibilidade de poder selecionar uma linha de EBV ajuda na visualização dos dados, principalmente ao indicar os casais que se destacaram naquela característica. Acredito que no futuro isso possibilitará ter um gradiente de seleção com base nesses dados e";
- "Acredito que poderia ser apresentado por outro tipo de gráfico, talvez um pouco mais simplificado, pensando num público mais leigo como os produtores.";

Figura 22 – Resultado da Análise das Cores Utilizadas no Gráfico de Cordas



Fonte — Produzida pelo próprio Autor

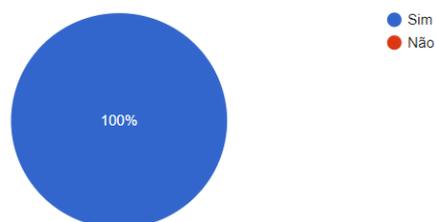
Comentários que os pesquisadores deixaram como contribuição:

- "Poderia melhor a escala ou as cores. Acho que poderiam ser feitos mais testes para que fique visualmente mais fácil identificar os melhores, e que seja mais intuitivo, do que sempre se recorrer a legenda";
- "Sim, Com a legenda ficou ainda mais fácil a visualização";
- "As cores utilizadas no gráfico de cordas auxiliam na análise porém acho que machos e fêmeas parentais também deveriam estar diferenciados por cores";
- "As cores utilizadas se misturam e fica difícil enxergar";
- "Com as legendas ficou mais fácil compreender sobre as cores quentes e frias, mas acredito que algo mais específico é auto explicativo poderia ser usado, por exemplo se essa ferramenta for usada posteriormente por pessoas leigas (produtores, por exemplo) acredito que possa ocorrer uma dificuldade em compreender a ferramenta";
- "Facilita na visualização dos quadros abaixo do gráfico de cordas. No entanto, pode gerar dúvidas para o produtor, no caso do verde sinalizar algo bom e vermelho proibição e";
- "Acho que ficaria mais chamativo associar cores como azul ou verde para famílias que apresentam uma quantidade maior , e cores como laranja e vermelho para quantidade mínima. Pensando nos produtores, acredito que seria mais adequado.".

Figura 23 – Resultado da Análise da Utilização da Legenda

No gráfico de cordas a legenda com as cores e o valores que foram utilizados para classificação dos peixes auxilia na identificação das melhores características ?

10 respostas

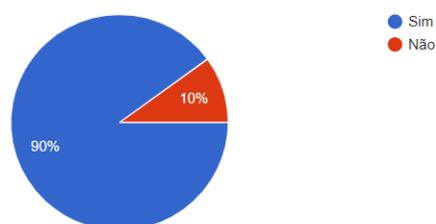


Fonte — Produzida pelo próprio Autor

Figura 24 – Resultado da Análise sobre as Características

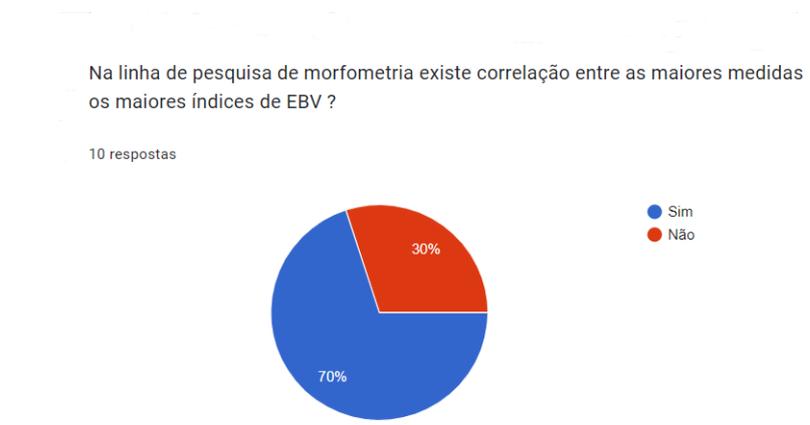
Na linha de pesquisa de morfometria, as características demonstradas no filtro tipo de dados auxiliam na análise?

10 respostas



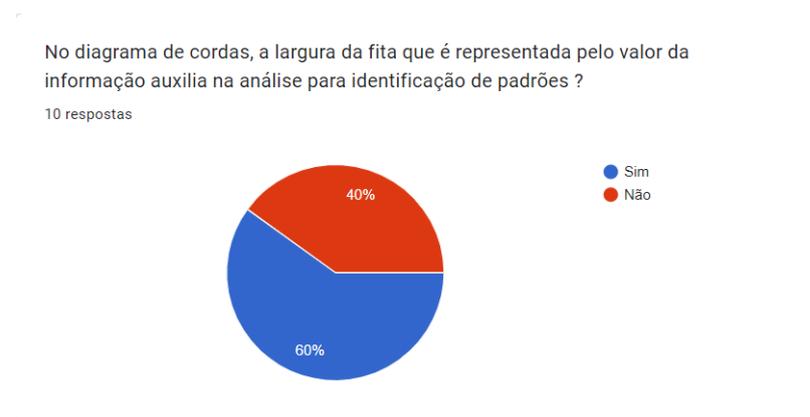
Fonte — Produzida pelo próprio Autor

Figura 25 – Resultado da Análise sobre as Medidas e os EBV



Fonte — Produzida pelo próprio Autor

Figura 26 – Resultado da Análise para Identificação de Padrões

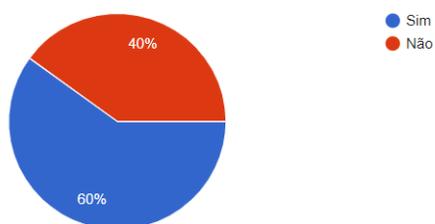


Fonte — Produzida pelo próprio Autor

Figura 27 – Resultado da Análise das Técnicas de Visualização Utilizadas

O conjunto das duas visualizações auxilia na análise para identificação da reprodução dos melhores casais ?

10 respostas

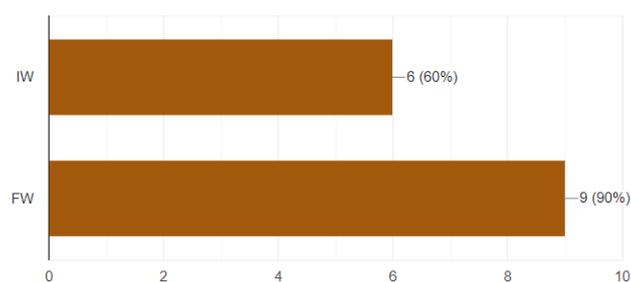


Fonte — Produzida pelo próprio Autor

Figura 28 – Resultado da Análise das Características na Área de Pesquisa Resistência

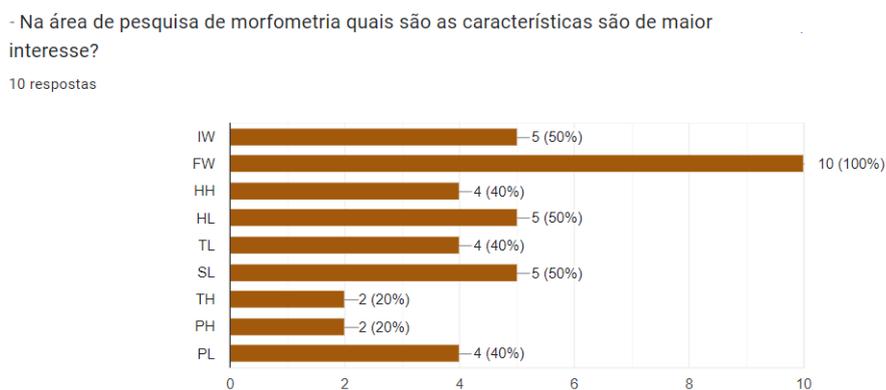
Na área de pesquisa de resistência quais são as características são de maior interesse?

10 respostas



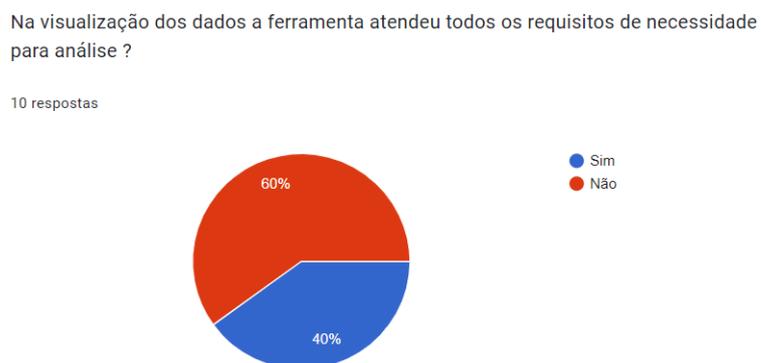
Fonte — Produzida pelo próprio Autor

Figura 29 – Resultado da Análise das Características na Área de Pesquisa Morfometria



Fonte — Produzida pelo próprio Autor

Figura 30 – Resultado sobre a Análise Geral

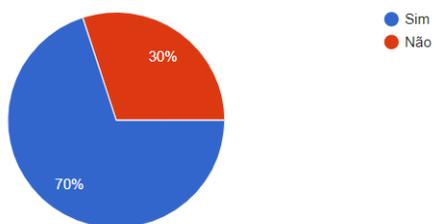


Fonte — Produzida pelo próprio Autor

Figura 31 – Resultado da Análise da Correlação entre as Características

Na análise visual foi possível identificar alguma correlação entre as características independente dos padrões que são demonstrados?

10 respostas



Fonte — Produzida pelo próprio Autor

7 Conclusão

7.1 Observações Finais

Nesse capítulo será detalhado sobre a conclusão alcançada com o uso da ferramenta desenvolvida. Baseado nos resultados obtidos através das respostas dos pesquisadores demonstrou-se a viabilidade do uso da ferramenta para análise. O uso de tecnologia web auxiliou no acesso a ferramenta. Os questionamentos efetuados aos pesquisadores sobre os resultados das análises efetuadas a partir da ferramenta mesmo não sendo unânime alcançaram índices positivos em sua maioria. No questionamento da usabilidade na utilização dos filtros os índices obtidos também foram positivos, a aceitação do software como ferramenta de análise foi positiva. As técnicas de visualização utilizadas para visualização atenderam as expectativas dos pesquisadores e conseguiram observar uma correlação entre os dados que foram demonstrados por elas. As lista de dados dos peixes ficaram muito próximas das planilhas utilizadas anteriormente facilitando o entendimento da área de aquicultura.

7.2 Contribuições do Trabalho

A utilização do Diagrama de Cordas como Técnica de Visualização demonstrando os casais dos peixes com o recurso da legenda classificando numa escala de cores quentes e frias, separando as medidas das menores até o maior valor da característica selecionada proporcionou aos pesquisadores uma análise das progenies. O uso das Coordenadas Paralelas como Técnica de Visualização demonstrou a variação dos índices genéticos(EBV), esse dado é uma informação muito importante para os pesquisadores na tomada de decisão. Para concluir, a união das duas técnicas num dashboard trouxe o recurso de uma análise mais abrangente, o pesquisador consegue com a seleção dos casais nas Coordenadas Paralelas comparar com as características de morfometria utilizada no Diagrama de Cordas.

7.3 Trabalhos Futuros

Para o desenvolvimento de um trabalho futuro poderia estudar outras Técnicas de Visualização que incorporasse no Dashboard e pudesse auxiliar na análise dos dados. Poderia estudar a viabilidade de trabalhar com mais cores e classificação da morfometria de forma variável com algum recursos de configuração.

7.4 Participações em Workshop e Publicações

Utilização das Técnicas de Visualização da Informação para Gestão Genética de Peixes. In: X Workshop do Programa de Pós-Graduação em Ciência da Computação da UNESP. Presidente Prudente, SP, Brasil: WPPGCC, 2021.

IV2022 - 26th International Conference on Information Visualisation Using Information Visualization Techniques for Fish Genetic Management Requena, Adriano; Feitosa, Juliana; de Camargo, Luiz Felipe; Brega, Jose Remo Ferreira, Viena Austria.

Referências

- Aouinti, F. et al. Satellite image restoration by applying the genetic approach to the wiener deconvolution. In: *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*. [S.l.: s.n.], 2016. p. 57–61. Citado na página 79.
- Astrinaki, M. et al. Zazz: Variant annotation and exploration of next generation sequencing variants. In: *2019 IEEE 19th International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.: s.n.], 2019. p. 856–860. ISSN 2471-7819. Citado na página 73.
- Aupetit, M. et al. A design study to identify inconsistencies in kinship information: The case of the 1000 genomes project. In: *2016 IEEE Pacific Visualization Symposium (PacificVis)*. [S.l.: s.n.], 2016. p. 254–258. ISSN 2165-8773. Citado 2 vezes nas páginas 16 e 71.
- Ayllón-Benítez, A. et al. Deciphering gene sets annotations with ontology based visualization. In: *2017 21st International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2017. p. 170–175. ISSN 2375-0138. Citado na página 73.
- Brunker, A. et al. Two-dimensional immersive cohort analysis supporting personalised medical treatment. In: *2019 23rd International Conference in Information Visualization – Part II*. [S.l.: s.n.], 2019. p. 34–41. Citado na página 75.
- Cao, Q. et al. Sgdb: A sports gene database for visualization of sports effects on human skeletal muscle gene expression. *IEEE Access*, v. 8, p. 20557–20562, 2020. ISSN 2169-3536. Citado na página 76.
- Chakraborty, R.; Hasija, Y. miderma: An integrated database and tool for analysis of mirnas associated with dermatological disorders. In: *2018 International Conference on Bioinformatics and Systems Biology (BSB)*. [S.l.: s.n.], 2018. p. 170–173. Citado na página 74.
- Coto, J. C.; Siles, F.; Mora-Rodríguez, R. Biocomputing platform module for cancer genomics and chemotherapy. In: *2016 IEEE 36th Central American and Panama Convention (CONCAPAN XXXVI)*. [S.l.: s.n.], 2016. p. 1–6. Citado na página 72.
- Cruz, A.; P. Arrais, J.; Machado, P. Interactive network visualization of gene expression time-series data. In: *2018 22nd International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2018. p. 574–580. ISSN 2375-0138. Citado na página 76.
- DOLSON, E.; OFRIA, C. Visualizing the tape of life: Exploring evolutionary history with virtual reality. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2018. (GECCO '18), p. 1553–1559. ISBN 9781450357647. Disponível em: <<https://doi.org/10.1145/3205651.3208301>>. Citado na página 79.
- EFTIMOV, T. et al. Performviz: A machine learning approach to visualize and understand the performance of single-objective optimization algorithms. In: *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA:

Association for Computing Machinery, 2020. (GECCO '20), p. 7–8. ISBN 9781450371278. Disponível em: <<https://doi.org/10.1145/3377929.3398182>>. Citado na página 80.

FABBRI, S. et al. Improvements in the StArt tool to better support the systematic review process. In: *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering - EASE '16*. New York, New York, USA: ACM Press, 2016. p. 1–5. ISBN 9781450336918. Disponível em: <<http://dl.acm.org/citation.cfm?doid=2915970.2916013>>. Citado na página 27.

FILIPIC, B.; TUSAR, T. A taxonomy of methods for visualizing pareto front approximations. In: *Proceedings of the Genetic and Evolutionary Computation Conference*. New York, NY, USA: Association for Computing Machinery, 2018. (GECCO '18), p. 649–656. ISBN 9781450356183. Disponível em: <<https://doi.org/10.1145/3205455.3205607>>. Citado na página 79.

GJEDREM T., R. M. Selection response in fish and shellfish: a review. *rev. aquacult.* 10 (1), 168–179. In: . [s.n.], 2018. Disponível em: <<https://doi.org/10.1111/raq.12154>>. Citado na página 17.

Glueck, M. et al. Phenostacks: Cross-sectional cohort phenotype comparison visualizations. *IEEE Transactions on Visualization and Computer Graphics*, v. 23, n. 1, p. 191–200, Jan 2017. ISSN 1941-0506. Citado na página 75.

Glueck, M. et al. Phenoblocks: Phenotype comparison visualizations. *IEEE Transactions on Visualization and Computer Graphics*, v. 22, n. 1, p. 101–110, Jan 2016. ISSN 1941-0506. Citado na página 13.

Glueck, M. et al. Phenoblocks: Phenotype comparison visualizations. *IEEE Transactions on Visualization and Computer Graphics*, v. 22, n. 1, p. 101–110, Jan 2016. ISSN 1941-0506. Citado na página 75.

Glueck, M. et al. Phenolines: Phenotype comparison visualizations for disease subtyping via topic models. *IEEE Transactions on Visualization and Computer Graphics*, v. 24, n. 1, p. 371–381, Jan 2018. ISSN 1941-0506. Citado na página 75.

GRIFFITHS A. J. F., W. S. R. L. R. C. C. S. B. *Introdução à genética*. [S.l.: s.n.], 2009. Citado na página 15.

Gupta, S. et al. Cancer-gene commonality network based on efficient retrieval methods. In: *2019 International Conference on Computing, Power and Communication Technologies (GUCON)*. [S.l.: s.n.], 2019. p. 463–468. Citado na página 72.

INSELBERG, A. The plane with parallel coordinates. *The Visual Computer*, v. 1, n. 2, p. 69–91, 1985. ISSN 1432-2315. Disponível em: <<https://doi.org/10.1007/BF01898350>>. Citado na página 45.

JAVASCRIPT. In: . [s.n.], 2021. Disponível em: <<https://developer.mozilla.org/pt-BR/docs/Web/JavaScript>>. Citado na página 39.

Jiang, Y. et al. Integrating gene expression data and pathway knowledge for in silico hypothesis generation with impres. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.: s.n.], 2018. p. 102–107. Citado na página 74.

- KANZKI, B. S. et al. Goat: Genetic output analysis tool: An open source gwas and genomic region visualization tool. In: *Proceedings of the 6th International Conference on Digital Health Conference*. New York, NY, USA: Association for Computing Machinery, 2016. (DH '16), p. 55–59. ISBN 9781450342247. Disponível em: <<https://doi.org/10.1145/2896338.2897729>>. Citado na página 77.
- KAUSE A., M. H. M. K. B. A. Genotype-by-environment interaction of growth traits in rainbow trout (*oncorhynchus mykiss*): a continental scale study.. *j. anim. sci.* 91 (12), 5572–5581. In: . [s.n.], 2013. Disponível em: <<https://doi.org/10.2527/jas2012-5949>>. Citado na página 17.
- KELLING W. M. HOCHACHKA, D. F. M. R. R. C. G. S.; HOOKER, G. *Data-intensive Science: A New Paradigm for Biodiversity Studies*. [S.l.: s.n.], 2009. Citado na página 16.
- Khwaja, M.; Kalofonou, M.; Toumazou, C. A deep belief network system for prediction of dna methylation. In: *2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*. [S.l.: s.n.], 2017. p. 1–4. Citado 2 vezes nas páginas 13 e 71.
- Kilgore, P. C. S. R. et al. Improving comprehension of large taxonomic graphs. In: *2019 23rd International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2019. p. 398–404. ISSN 2375-0138. Citado na página 73.
- KIRK, A. *Data Visualization: A Successful Design Process*. Packt Pub., 2012. (Community experience distilled). ISBN 9781849693479. Disponível em: <<https://books.google.com.br/books?id=I4qBVLfd3t4C>>. Citado 2 vezes nas páginas 17 e 19.
- KIRK, A. *Data Visualization: A Successful Design Process: A Structured Design Approach To Equip You With The Knowledge Of How To Successfully Accomplish Any Data Visualization Challenge Efficiently And Effectively*. [S.l.]: Packt Pub, 2012. 189 p. ISBN 9781849693462 9781849693462. Citado 2 vezes nas páginas 44 e 45.
- KUZNETSOVA, I. et al. Circularized visualisation of genetic interactions. In: *Proceedings of the 26th International Conference on World Wide Web Companion*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017. (WWW '17 Companion), p. 225–226. ISBN 9781450349147. Disponível em: <<https://doi.org/10.1145/3041021.3058587>>. Citado na página 77.
- Lee, D. J. L. et al. You can't always sketch what you want: Understanding sensemaking in visual query systems. *IEEE Transactions on Visualization and Computer Graphics*, v. 26, n. 1, p. 1267–1277, Jan 2020. ISSN 1941-0506. Citado na página 76.
- Lee, S.; Lee, D. Visualizing multifunctional ppi network with gene ontology annotation. In: *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.: s.n.], 2019. p. 2069–2074. Citado na página 79.
- LORENZO, A. D. et al. An analysis of dimensionality reduction techniques for visualizing evolution. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2019. (GECCO '19), p. 1864–1872. ISBN 9781450367486. Disponível em: <<https://doi.org/10.1145/3319619.3326868>>. Citado na página 80.

- Mazrouee, S.; Wang, W. Polycluster: Minimum fragment disagreement clustering for polyploid phasing. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, v. 17, n. 1, p. 264–277, Jan 2020. ISSN 1557-9964. Citado na página 79.
- MCPHEE, N. F. et al. Visualizing genetic programming ancestries. In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2016. (GECCO '16 Companion), p. 1419–1426. ISBN 9781450343237. Disponível em: <<https://doi.org/10.1145/2908961.2931741>>. Citado na página 78.
- MCPHEE, N. F. et al. Visualizing genetic programming ancestries using graph databases. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2017. (GECCO '17), p. 245–246. ISBN 9781450349390. Disponível em: <<https://doi.org/10.1145/3067695.3075617>>. Citado na página 80.
- MEDVET, E.; TUVSAR, T. The du map: A visualization to gain insights into genotype-phenotype mapping and diversity. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2017. (GECCO '17), p. 1705–1712. ISBN 9781450349390. Disponível em: <<https://doi.org/10.1145/3067695.3082554>>. Citado na página 77.
- MEEKS. *D3.js in Action: Data Visualization with JavaScript*. [S.l.: s.n.], 2018. ISBN 978-1-617-29448-8. Citado na página 39.
- MICHALAK, K. Low-dimensional euclidean embedding for visualization of search spaces in combinatorial optimization. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2019. (GECCO '19), p. 27–28. ISBN 9781450367486. Disponível em: <<https://doi.org/10.1145/3319619.3326761>>. Citado na página 79.
- MICROSOFT. In: . [s.n.], 2021. Disponível em: <<https://docs.microsoft.com/pt-br/dotnet/csharp/tour-of-csharp/>>. Citado na página 38.
- MUNZNER, T. *Visualization Analysis Design*. [S.l.: s.n.], 2014. ISBN -13: 978-1466508910. Citado na página 22.
- Nobre, C. et al. Lineage: Visualizing multivariate clinical data in genealogy graphs. *IEEE Transactions on Visualization and Computer Graphics*, v. 25, n. 3, p. 1543–1558, March 2019. ISSN 1941-0506. Citado na página 74.
- Palleschi, A.; Crielesi, M. A visual analytics system of data gathered from colonial seabirds. In: *2019 23rd International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2019. p. 224–227. ISSN 2375-0138. Citado 2 vezes nas páginas 16 e 71.
- Pawar, S.; Stanam, A. Scalable, reliable and robust data mining infrastructures. In: *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*. [S.l.: s.n.], 2020. p. 123–125. Citado na página 79.
- PEARSON, K. Report on certain enteric fever inoculation statistics. *bmj*. In: . [S.l.: s.n.], 1904. Citado na página 23.

Perkins, M.; Daniels, K. Visualizing dynamic gene interactions to reverse engineer gene regulatory networks using topological data analysis. In: *2017 21st International Conference Information Visualisation (IV)*. [S.l.: s.n.], 2017. p. 384–389. ISSN 2375-0138. Citado na página 76.

RAUSCHMAYER, D. A. *Speaking JavaScript: An In-Depth Guide for Programmers*. O'Reilly Media, 2014. Disponível em: <<https://http://speakingjs.com/es5/index.html>>. Citado na página 39.

Santos, V. C. A. et al. Metagenomics-based signature clustering and interactive visualization analysis. In: *2018 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2018. p. 1–8. ISSN 2161-4407. Citado na página 74.

Sawada, K. et al. Analyzing the mutation frequencies and correlation of genetic diseases in worldwide populations using big data processing, clustering, and predictive analytics. In: *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*. [S.l.: s.n.], 2017. p. 1459–1464. Citado na página 72.

SEKANINA, L.; KAPUSTA, V. Visualisation and analysis of genetic records produced by cartesian genetic programming. In: *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2016. (GECCO '16 Companion), p. 1411–1418. ISBN 9781450343237. Disponível em: <<https://doi.org/10.1145/2908961.2931740>>. Citado na página 80.

Soukaina, B. et al. Optimization of the attribute vector by genetic approach: Application to the classification of characters. In: *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*. [S.l.: s.n.], 2016. p. 405–409. Citado na página 80.

SQLITE. In: . [s.n.], 2022. Disponível em: <<https://www.sqlite.org/index.html>>. Citado na página 39.

Stefanou, K. et al. An intelligent web-based system for the detection and visualization of biomarkers in microdeletion and microduplication syndromes. In: *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.: s.n.], 2020. p. 217–221. ISSN 2471-7819. Citado na página 73.

Thierbach, K. et al. Biology-inspired visualization of morphogenetic motion in the zebrafish endoderm. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2016. p. 925–929. ISSN 1945-8452. Citado na página 13.

Thierbach, K. et al. Biology-inspired visualization of morphogenetic motion in the zebrafish endoderm. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. [S.l.: s.n.], 2016. p. 925–929. ISSN 1945-8452. Citado na página 72.

Vinh Nguyen, Q. et al. A mobile tool for interactive visualisation of genomics data. In: *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. [S.l.: s.n.], 2018. p. 688–697. ISSN 2474-3828. Citado na página 71.

Vrahatis, A. G. et al. Visualizing high-dimensional single-cell rna-seq data via random projections and geodesic distances. In: *2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. [S.l.: s.n.], 2019. p. 1–6. Citado na página 79.

Wang, N. et al. Analyzing the canopy light distribution among different poplar genotypes using terrestrial laser scanner and the greenlab model. In: *2016 IEEE International Conference on Functional-Structural Plant Growth Modeling, Simulation, Visualization and Applications (FSPMA)*. [S.l.: s.n.], 2016. p. 214–223. Citado na página 72.

WANG, R.; CLUNE, J.; STANLEY, K. O. Vine: An open source interactive data visualization tool for neuroevolution. In: *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York, NY, USA: Association for Computing Machinery, 2018. (GECCO '18), p. 1562–1564. ISBN 9781450357647. Disponível em: <<https://doi.org/10.1145/3205651.3208236>>. Citado na página 80.

Zeng, S. et al. Knowledge base commons (kbcommons) v1.0: A multi omics' web-based data integration framework for biological discoveries. In: *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. [S.l.: s.n.], 2018. p. 589–594. Citado na página 74.

A Dados dos estudos incluídos

Neste apêndice são apresentados os dados extraídos dos estudos incluídos na revisão, através da Tabela 9.

Tabela 9 – Artigos incluídos na revisão.

Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
A deep belief network system for prediction of DNA methylation	Khwaja, Kalofonou e Toumazou (2017)	M. Khwaja and M. Kalofonou and C. Toumazou	Estudo para identificação de Câncer	Gráfico de Barras	Banco de Dados	Não Mencionada
A design study to identify inconsistencies in kinship information: The case of the 1000 Genomes project	Aupetit et al. (2016)	M. Aupetit and E. Ullah and R. Rawi and H. Bensmail	Estudo para examina variantes genéticas em diferentes indivíduos para detectar variantes associadas a específicas	Gráfico de barras e dispersão	Não Mencionada	Não Mencionada
A Mobile Tool for Interactive Visualisation of Genomics Data	Vinh Nguyen et al. (2018)	Q. Vinh Nguyen and Z. Qu and M. Lin Huang and C. Wei Lau and S. Simoff and D. R. Catchpoole	Estudos para descoberta de leucemia linfoblástica em crianças	Mapa de Calor	Banco de Dados	Linguagens para android e ios unity3D
A Visual Analytics System of Data Gathered from Colonial Seabirds	Palleschi e Crielesi (2019)	A. Palleschi and M. Crielesi	Genética sobre a criação de ave marinha cagarra de Scopoli	grafico de dispersao, mapas e treemap	Banco de Dados	Não Mencionada

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Analyzing the canopy light distribution among different poplar genotypes using terrestrial laser scanner and the GreenLab model	Wang et al. (2016)	N. Wang and Y. Shen and J. Hua and H. Wang and M. Kang and X. Su	Genética sobre Plantação de Árvores	grafico de dispersao, grafico de linhas e barras	Não Mencionada	Não Mencionada
Analyzing the Mutation Frequencies and Correlation of Genetic Diseases in Worldwide Populations Using Big Data Processing, Clustering, and Predictive Analytics	Sawada et al. (2017)	K. Sawada and M. W. Clark and N. Alshurafa and M. Pourhomayoun	Estudo de Genética sobre o risco de desenvolvimento das doenças osteoporose e cardiovascular	Gráfico de Barras e Histogramas	Arquivos CSV	python e Biblioteca Matplotlib
Biocomputing platform module for cancer genomics and chemotherapy	Coto, Siles e Mora-Rodríguez (2016)	J. C. Coto and F. Siles and R. Mora-Rodríguez	Estudo de Genética para identificação de tratamento com a quimioterapia	Gráfico de Dispersão	Banco de Dados	linguagem R e Phyton
Biology-inspired visualization of morphogenetic motion in the zebrafish endoderm	Thierbach et al. (2016b)	K. Thierbach and G. Shah and I. Glauche and J. Huisken and I. Roeder and N. Scherf	Genética de peixe zebra	Gráfico de Linha	Não Mencionada	Não Mencionada
Cancer-Gene Commonality Network based on Efficient Retrieval Methods	Gupta et al. (2019)	S. Gupta and H. Sharma and Y. Sharma and A. Purwar	Identificacao do desenvolvimento do câncer através da genetica	Gráfico Interligado	Banco de Dados	Não Mencionada

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Zazz: Variant Annotation and Exploration of Next Generation Sequencing Variants	Astrinaki et al. (2019)	M. Astrinaki and A. Kanterakis and H. Latsoudis and G. Potamias and D. Kafetzopoulos	Examinar padrão de herança e analisar fenótipos complexos	Gráfico de Barras	Banco de Dados e Json	Phyton e dc.js
An intelligent web-based system for the detection and visualization of biomarkers in Microdeletion and Microduplication Syndromes	Stefanou et al. (2020)	K. Stefanou and C. Bellos and G. Stergios and A. Fyraridis and P. Ladias and P. Sakaloglou and C. Kostoulas and S. Markoula and I. Georgiou	Estudo Genético para identificação de retardo mental e informações para possíveis medicamentos	Gráficos de Barras e Histogramas	Banco de Dados	CSharp, ASP.NET e Genovese
Deciphering Gene Sets Annotations with Ontology Based Visualization	Ayllón-Benítez et al. (2017)	A. Ayllón-Benítez and P. Thébault and J. T. Fernández-Breis and M. Quesada-Martínez and F. Mougin and R. Bourqui	Estudo Genético para identificar os principais reguladores que gerenciam a defesa imunológica	Treemap	Banco de dados	Não Mencionada
Improving Comprehension of Large Taxonomic Graphs	Kilgore et al. (2019)	P. C. S. R. Kilgore and M. Trutschl and U. Cvek and J. S. Alexander	Estudo Genético para identificar a ancestralidade através das fezes do macaco	Grafos	Banco de Dados	Não Mencionada

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Integrating Gene Expression Data and Pathway Knowledge for In Silico Hypothesis Generation with IMPRes	Jiang et al. (2018)	Y. Jiang and D. Wang and D. Xu and T. Joshi	Genética Celular	Gráficos Interligado em Redes	Não Mencionada	Não Mencionada
Knowledge Base Commons (KBCommons) v1.0: A multi OMICS' web-based data integration framework for biological discoveries	Zeng et al. (2018)	S. Zeng and Z. Lyu and S. R. K. Narisetti and D. Xu and T. Joshi	Análise de geral de genética de plantas, humanos e animais.	gráfico de dispersão, mapas de calor e agrupamento hierárquico	Mysql e Mongoddb	Python, Php, Javascript, Html e Tecnologias como Bootstrap, Highcharts e Google Charts
Lineage: Visualizing Multivariate Clinical Data in Genealogy Graphs	Nobre et al. (2019)	C. Nobre and N. Gehlenborg and H. Coon and A. Lex	Estudo de Fatores Genéticos e ambientais em Suicídios	Árvore Multivariada	Banco de dados	Typescript e D3
Metagenomics-based signature clustering and interactive visualization analysis	Santos et al. (2018)	V. C. A. Santos and L. Correa and B. Meiguins and G. Oliveira and R. Alves	Estudo da presença de microrganismos em ambientes	bubble Tree	Não Mencionada	java Script
miDerma: An Integrated Database and Tool for Analysis of miRNAs associated with Dermatological Disorders	Chakraborty e Hasija (2018)	PR. Chakraborty and Y. Hasija	Estudo de genes associados a doenças de pele	Visualização em redes	Banco de Dados	Python, Java Script, Html e CSS

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
PhenoBlocks: Phenotype Comparison Visualizations	Glueck et al. (2016b)	M. Glueck and P. Hamilton and F. Chevalier and S. Breslav and A. Khan and D. Wigdor	Estudo Genético para diagnóstico diferencial de doenças genéticas raras	Sunburst	Banco de dados e Json	Não Mencionada
PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models	Glueck et al. (2018)	M. Glueck and M. P. Naeni and F. Doshi-Velez and F. Chevalier and A. Khan and D. Wigdor and M. Brudno	Estudo Genético para apoiar a interpretação de subtipos de doenças	Gráfico de dispersão, Hierarquia Radial, Gráfico de linha do tempo e Gráfico de barras	Não Mencionada	Não Mencionada
PhenoStacks: Cross-Sectional Cohort Phenotype Comparison Visualizations	Glueck et al. (2017)	M. Glueck and A. Gvozdk and F. Chevalier and A. Khan and M. Brudno and D. Wigdor	Estudo Genético para Diagnóstico de Doenças	Sunburst, Gráfico de Dispersão	banco de dados	Phyton, D3 e Bibliotecas SciPy
Two-Dimensional Immersive Cohort Analysis Supporting Personalised Medical Treatment	Brunker et al. (2019)	A. Brunker and D. Catchpoole and P. Kennedy and S. Simoff and Q. V. Nguyen	Estudo Genético sobre semelhança de 2 tipos de câncer (Leucemia Linfoblástica e Rabdomiossarcoma)	Mapa de Calor	Não Mencionada	Não Mencionada

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Visualizing Dynamic Gene Interactions to Reverse Engineer Gene Regulatory Networks Using Topological Data Analysis	Perkins e Daniels (2017)	M. Perkins and K. Daniels	Melhoria da Dinâmica do Genoma	Visualização 3d Espaço Temporal	Não Mencionada	C++
SGDB: A Sports Gene Database for Visualization of Sports Effects on Human Skeletal Muscle Gene Expression	Cao et al. (2020)	Q. Cao and Z. Deng and J. Liu and X. Li	Estudo Genético para analisar mudanças nos genes após diferentes formas de exercícios e suas relações com o ser humano, gênero e idade	Gráfico de Barras	Banco de dados e Mysql	Php
You can't always sketch what you want: Understanding Sensemaking in Visual Query Systems	Lee et al. (2020)	D. J. -L. Lee and J. Lee and T. Siddiqui and J. Kim and K. Karahalios and A. Parameswaran	Estudar o Padrão de Genes	Gráfico de Linhas	Banco de dados	Phyton
Interactive Network Visualization of Gene Expression Time-Series Data	Cruz, P. Arrais e Machado (2018)	A. Cruz and J. P. Arrais and P. Machado	O comportamento resultante da invecção por HIV-I no gene	Rede Bidimensional	Banco de Dados	Não Mencionada

(continuação)						
Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Circularized Visualisation of Genetic Interactions	Kuznetsova et al. (2017)	Kuznetsova, Irina and Filipovska, Aleksandra and Rackham, Oliver and Lugmayr, Artur and Holzinger, Andreas	Esse estudo forneceu ao biólogo um mapa de interações genéticas, o qual é essencial para tirar conclusões sobre mudanças nas interações de genes em doenças	Diagrama de Círculos	Não Mencionada	Flash
The DU Map: A Visualization to Gain Insights into Genotype-Phenotype Mapping and Diversity	Medvet e Tuvsar (2017)	Medvet, Eric and Tuvsar, Tea	Diversidade entre o mapa do genotipo e o fenotipo	Mapa de Calor e Gráfico de linha	Não Mencionada	Não Mencionada
GOAT: Genetic Output Analysis Tool: An Open Source GWAS and Genomic Region Visualization Tool	Kanzki et al. (2016)	Kanzki, Beatriz S. and Dupuy, Victor and Urvoy, Cedric and Belghait, Fodil and April, Alain and Harvey, Francois and Marois-Blanchet, Francois-Christophe and Phillips, Michael S. and Tremblay, Johanne and Hamet, Pavel	Estudo Genético para identificacao de doenças diabetes e doenças renais crônicas	Visualização 3D	Mysql	Python e JavaScript

(continuação)

Título	Referência	Autor	Tipos de Genética	Técnicas de visualização	Formato dos dados	Linguagens utilizadas
Visualizing Genetic Programming Ancestries	McPhee et al. (2016)	McPhee, Nicholas Freitag and Casale, Maggie M. and Finzel, Mitchell and Helmuth, Thomas and Spector, Lee	Hepatite C	Gráfico de Pontos	Banco de Dados	Clojush

Fonte - Produzida pelo autor.

B Dados dos estudos excluídos

Neste apêndice são apresentados os estudos excluídos na revisão, através da Tabela 10.

Tabela 10 – Artigos excluídos na revisão.

Título	Referência	Autor	Motivo
Satellite Image Restoration by Applying the Genetic Approach to the Wiener Deconvolution	Aouinti et al. (2016)	F. Aouinti and M. Nasri and M. Moussaoui and S. Benchaou and K. Zinedine	Não possui foco com os temas desejados
Scalable, Reliable and Robust Data Mining Infrastructures OK	Pawar e Stanam (2020)	S. Pawar and A. Stanam	Não possui foco com os temas desejados
Visualizing multifunctional PPI network with Gene Ontology annotation	Lee e Lee (2019)	S. Lee and D. Lee	Não possui foco com os temas desejados
Visualizing High-Dimensional Single-Cell RNA-seq Data via Random Projections and Geodesic Distances	Vrahatis et al. (2019)	A. G. Vrahatis and S. K. Tasoulis and G. N. Dimitrakopoulos and V. P. Plagianakos	Não possui foco com os temas desejados
PolyCluster: Minimum Fragment Disagreement Clustering for Polyploid Phasing	Mazrouee e Wang (2020)	S. Mazrouee and W. Wang	Não possui foco com os temas desejados
A Taxonomy of Methods for Visualizing Pareto Front Approximations	Filipic e Tusar (2018)	Filipic, Bogdan and Tusar, Tea	Não possui foco com os temas desejados
Visualizing the Tape of Life: Exploring Evolutionary History with Virtual Reality	Dolson e Ofria (2018)	Dolson, Emily and Ofria, Charles	Não possui foco com os temas desejados
Visualizing Swarm Behavior with a Particle Density Map	Lee e Lee (2019)	Lee, Hyeon-Chang and Kim, Yong-Hyuk	Não possui foco com os temas desejados
Low-Dimensional Euclidean Embedding for Visualization of Search Spaces in Combinatorial Optimization	Michalak (2019)	Michalak, Krzysztof	Não possui foco com os temas desejados

(continuação)

Título	Referência	Autor	Motivo
Optimization of the Attribute Vector by Genetic Approach: Application to the Classification of Characters	Soukaina et al. (2016)	B. Soukaina and M. Nasri and F. Aouinti and K. Zinedine and O. El Melhaoui	Não possui foco com os temas desejados
PerformViz: A Machine Learning Approach to Visualize and Understand the Performance of Single-Objective Optimization Algorithms	Eftimov et al. (2020)	Eftimov, Tome and Hribar, Rok and Škvorc, Urban and Popovski, Gorjan and Petelin, Gašper and Korošec, Peter	Não possui foco com os temas desejados
Visualisation and Analysis of Genetic Records Produced by Cartesian Genetic Programming	Sekanina e Kapusta (2016)	Sekanina, Lukas and Kapusta, Vlastimil	Não possui foco com os temas desejados
VINE: An Open Source Interactive Data Visualization Tool for Neuroevolution	Wang, Clune e Stanley (2018)	Wang, Rui and Clune, Jeff and Stanley, Kenneth O.	Não possui foco com os temas desejados
Visualizing Genetic Programming Ancestries Using Graph Databases	McPhee et al. (2017)	McPhee, Nicholas Freitag and Casale, Maggie M. and Finzel, Mitchell and Helmuth, Thomas and Spector, Lee	Não possui foco com os temas desejados
An Analysis of Dimensionality Reduction Techniques for Visualizing Evolution	Lorenzo et al. (2019)	De Lorenzo, Andrea and Medvet, Eric and Tusar, Tea and Bartoli, Alberto	Não possui foco com os temas desejados

Fonte - Produzida pelo autor.