



Universidade do Minho  
Escola de Engenharia



INSTITUTO DE CIÊNCIA E INOVAÇÃO  
PARA A BIO-SUSTENTABILIDADE

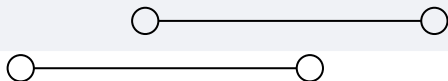


centro de **biologia**  
**molecular** e ambiental

# Development of a User-Friendly Graphical Platform for Molecular Characterization of Macroscopic Parasites in Fish



[https://github.com/JohnnyFarians24/Project\\_Bioinformatics](https://github.com/JohnnyFarians24/Project_Bioinformatics)



# Conteúdos

✦ 01 ✦

Introdução

✦ 02 ✦

Problema

✦ 03 ✦

Objetivos

✦ 04 ✦

Metodologia

✦ 05 ✦

Resultados esperados





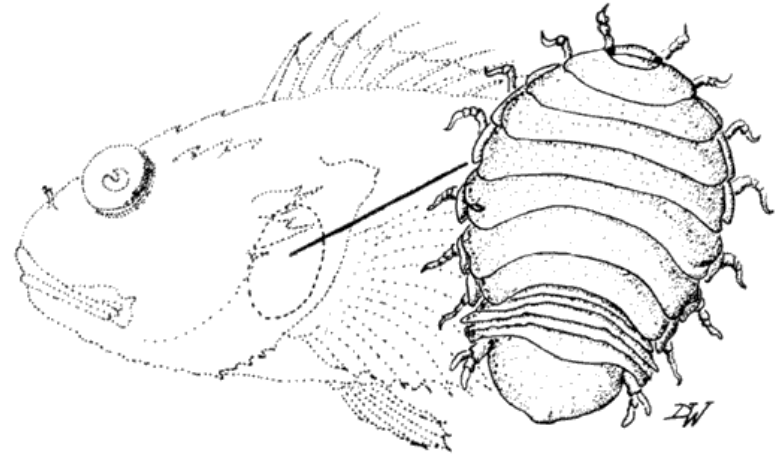
✦ 01 ✦

# Introdução



# Importância dos Parasitas em Sistemas Biológicos

- Parasitas são **organismos onnipresentes** e representam uma parte significativa da biodiversidade global.
- Encontram-se entre as formas de vida mais bem-sucedidas, estando **presentes em vários filos**.
- Em ambientes aquáticos, destacam-se pela sua **prevalência e sucesso evolutivo**, especialmente nas interações com peixes.



**Fig.1-** Exemplo de um isópode (*Isopoda*).  
<https://wildlife.ca.gov/Conservation/Marine/Parasites>

# Impacto dos Parasitas na Ecologia dos Peixes

- Parasitas são **capazes de alterar fatores ecológicos** e parâmetros usados como marcadores de saúde ou reprodução dos peixes.
- As mudanças induzidas no comportamento do hospedeiro **afetam:**
  - Distribuição e escolha de habitat
  - Composição da dieta
  - Comportamento sexual
- Essas alterações **impactam a ecologia dos peixes**, os seus predadores e presas.



**Fig.2-** Corvo-Marinho (*Phalacrocorax carbo*) a pescar.  
<https://www.pexels.com/pt-br/foto/corvo-marinho-capturando-peixes-em-aguas-oceanicas-30920841/>

# Metabarcoding e NGS: Revolução na Identificação de Parasitas Aquáticos



- Métodos precisos de identificação são cruciais para compreender o impacto dos parasitas.
- **Metabarcoding** tornou-se popular na última década para:
  - **Deteção de parasitas.**
  - **Amostragem não invasiva, rápida e exaustiva** da composição das comunidades de parasitas.



- **NGS** permite obter milhões de leituras simultaneamente, **facilitando:**
  - Mapeamento de espécies
  - Biomonitorização
  - Análise de conteúdo intestinal
  - Genómica populacional
- Metabarcoding de DNA permite a **deteção simultânea de múltiplas espécies**, mesmo em quantidades vestigiais.
- Método é **rentável, não invasivo** e reduz dependência de identificação morfológica.

# Desafios, Limitações e Avanços Recentes no Metabarcoding de Parasitas



- Necessidade de **bibliotecas de referência de alta qualidade** e atualizadas (ex: GenBank, BOLD).
- Dados erróneos nas bibliotecas podem causar erros de identificação.
- **Curadoria manual é difícil em ambientes marinhos** devido à grande diversidade.
- **Ferramentas bioinformáticas** especializadas **são essenciais** para a detecção eficiente de parasitas de peixes.



- **Metabarcoding de eDNA** tornou-se comum no estudo de peixes marinhos.
- Monitorização de pescarias com eDNA depende de conjuntos de dados exaustivos.
- Iniciadores universais (COI) podem amplificar espécies não alvo; **iniciadores específicos para peixes**, como **MiFish**, foram desenvolvidos para maior precisão.



✦ 02 ✦

# Problema





- Ferramentas como o **PMiFish**, desenvolvidas em Perl, melhoram a análise de eDNA de peixes usando **primers MiFish**, integrando **MEGA X** (árvores filogenéticas) e **USEARCH** (análise rápida de sequências).
- Limitações:
  - **Requer ambiente Linux**, dificultando o uso para diversos pesquisadores.
  - **Gestão de grandes volumes de dados** de sequenciação é **complexa**.
  - **Curadoria de bases de dados** de referência (ex: parasitas de peixes) é **trabalhosa**.
  - Integração de múltiplos módulos bioinformáticos em sistemas unificados ainda é um desafio.

rogotoh/**PMiFish**



1 Contributor 0 Issues 12 Stars 4 Forks



**Fig.3-** Ferramenta PMiFish criada pelo usuário rogotoh.  
<https://github.com/rogotoh/PMiFish>



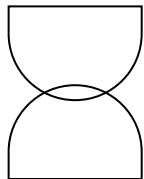
✦ 03 ✦

# Objetivos





- Desenvolver uma **interface gráfica multiplataforma**, intuitiva e fácil de usar **para análise de dados de metabarcoding**.
- Simplificar processos como:
  - Construção de árvores filogenéticas
  - Identificação de espécies de peixes e parasitas
  - Catalogação e fusão de amostras
- Permitir que **não especialistas realizem análises** moleculares complexas.
- **Enriquecer a biblioteca** de referência com sequências **de parasitas de peixes**.
- **Melhorar o pipeline PMiFish** com uma interface amigável.
- **Minimizar resíduos alimentares** de peixes, aprimorando a detecção e controlo de parasitas.
- Utilizar técnicas avançadas de bioinformática e NGS para análise rápida e precisa de grandes volumes de dados.





04

# Metodologia

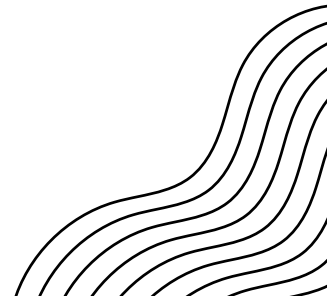
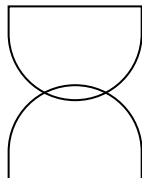




# Fluxo de Processamento de Dados de Sequenciação

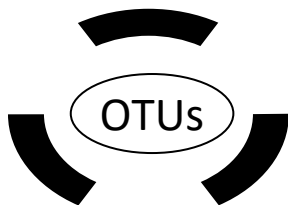


- Controlo de qualidade:
  - **FQC (baseado em FastQC)** para análise interativa de dados brutos. Permite verificar:
    - Padrões de abundância de leitura
    - Conteúdo de adaptadores
    - Distribuição de GC
    - Qualidade da sequência por base.
- Corte e limpeza:
  - **Trimmomatic** remove regiões de baixa qualidade e adaptadores.
- Fusão de leituras:
  - **PEAR** para leituras emparelhadas; FQC e Trimmomatic para leituras únicas.
- Análise paralela:
  - Suporte para dados emparelhados e únicos, garantindo filtragem consistente.

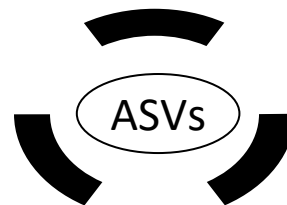




# Agrupamento e Denoising de Sequências



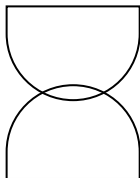
- Agrupam sequências com  **$\geq 97\%$  de identidade**, facilitando a análise de diversidade microbiana.
- Utilização do **VSEARCH**, ferramenta open-source, rápida e eficiente para **clustering de OTUs**, que substitui o USEARCH.



- Identificam **variantes únicas de sequência**, distinguindo diferenças de nucleótidos reais de erros de sequenciação.

Através da implementação do **DADA2** que possui:

- **Maior precisão.**
- **Mais ASVs** comparativamente a outras opções.
- **Melhor sensibilidade** na detecção de variantes.



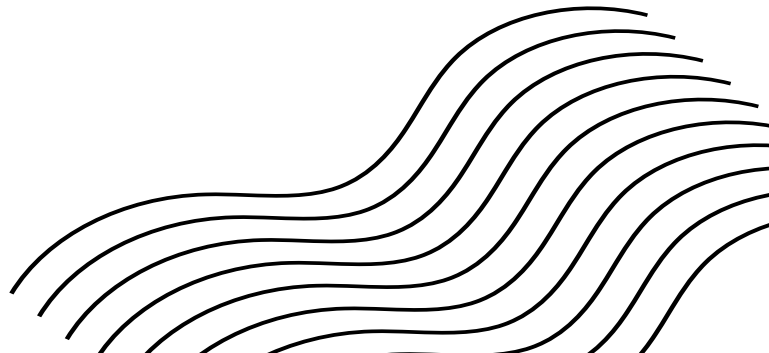
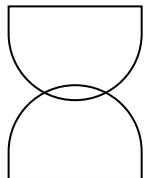


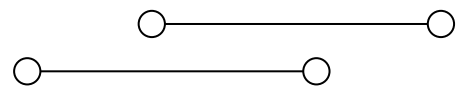
# Classificação Taxonômica Direta



Através da integração do **Kraken2**:

- Utiliza alinhamento de **k-mers** para classificar sequências de forma **rápida e precisa**.
- Suporta análise de amplicons e **metagenômica shotgun**.
- Classificação de rRNA 16S **em bases como SILVA**, Greengenes e RDP.





# Bases de Dados



## NCBI GenBank<sup>1</sup>

- **Maior base de dados de nucleótidos**, utilizada para referências cruzadas, mas com possíveis anotações erradas devido à submissão aberta.



## BOLD<sup>2</sup>

- Principal fonte para **genes COI**, com dados complementares ao GenBank; acesso automatizado **via BOLDigger** para contornar limites de consulta.



## SILVA e PR2<sup>3 4</sup>

- Bases de dados especializadas em rRNA **(16S, 18S)**, de alta qualidade, anotadas e atualizadas regularmente.

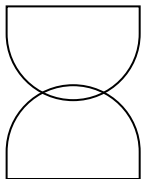
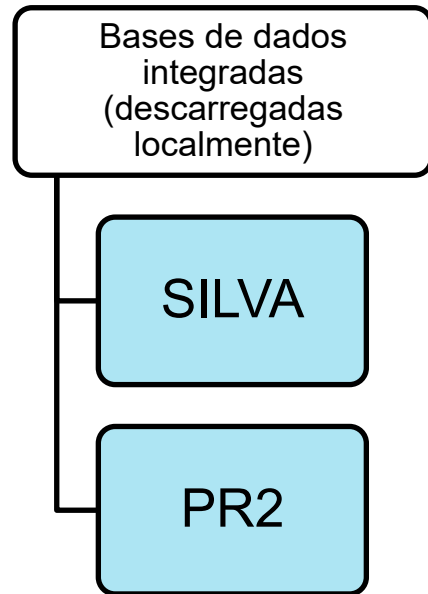
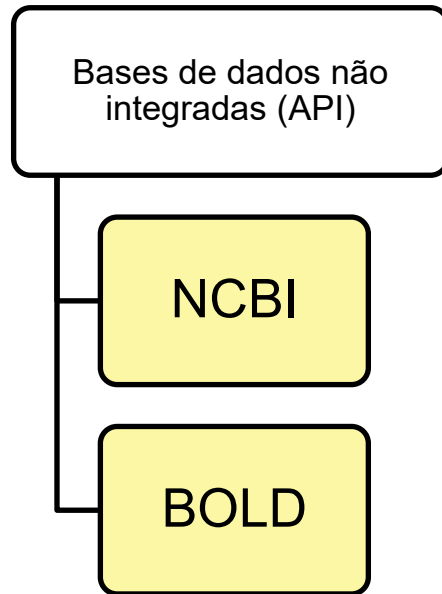


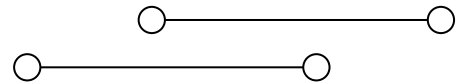


# Seleção e Integração Inteligente



- Utilizador pode **selecionar a base de dados** mais adequada para a sua análise.
- Incluindo selecionar a sua **própria base de dados** e fazer upload da mesma na plataforma.
- Com informações claras sobre **características e limitações de cada uma**.





# Análise Filogenética

## Alinhamento de Sequências

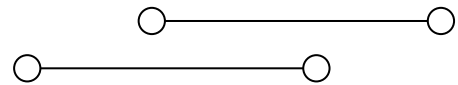


## Construção de Árvores Filogenéticas

Através da utilização do **MAFFT**:

- Reconhecido pela:
  - Precisão.
  - Rapidez.
  - Alinhamentos de alta qualidade.
- Essenciais para análises filogenéticas robustas.

- **FastTree 2:**
  - Topologia inicial da árvore filogenética:
    - Modelo GTR.
- **RAxML-NG:**
  - Otimizar o comprimento dos ramos:
    - Árvores mais precisas e informativas.
- O resultado são árvores que **ilustram as relações evolutivas** entre as espécies identificadas, **facilitando a interpretação biológica dos dados**.



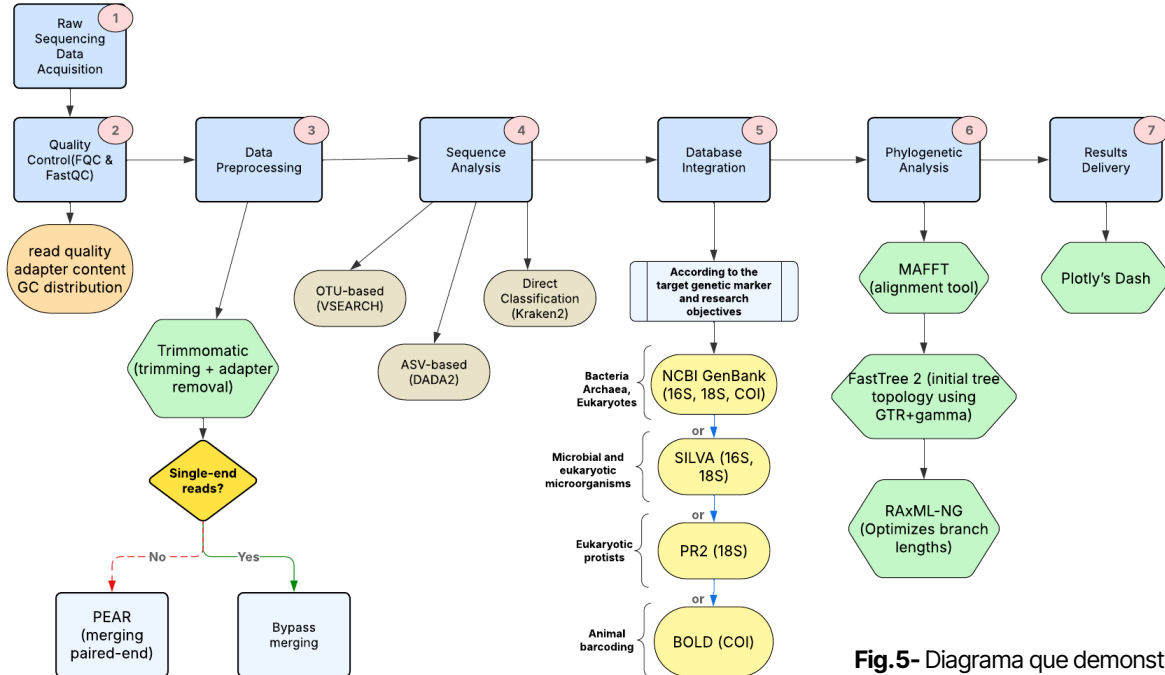
# Desenvolvimento da Interface de Utilizador

- Plataforma desenvolvida com **Plotly Dash**, que possui:
  - A capacidade da criação de **aplicações web** interativas e multiplataforma **em Python**.
  - Uma interface amigável, projetada para **simplificar fluxos de trabalho complexos**, tornando-os acessíveis a utilizadores não especialistas.
  - **Visualização dinâmica dos resultados**, incluindo gráficos, tabelas e árvores filogenéticas interativas.
  - A possibilidade da integração de todas as etapas do pipeline, num **ambiente unificado e de fácil utilização**.



**Fig.4-** Plotly Dash <https://dash.plotly.com/>

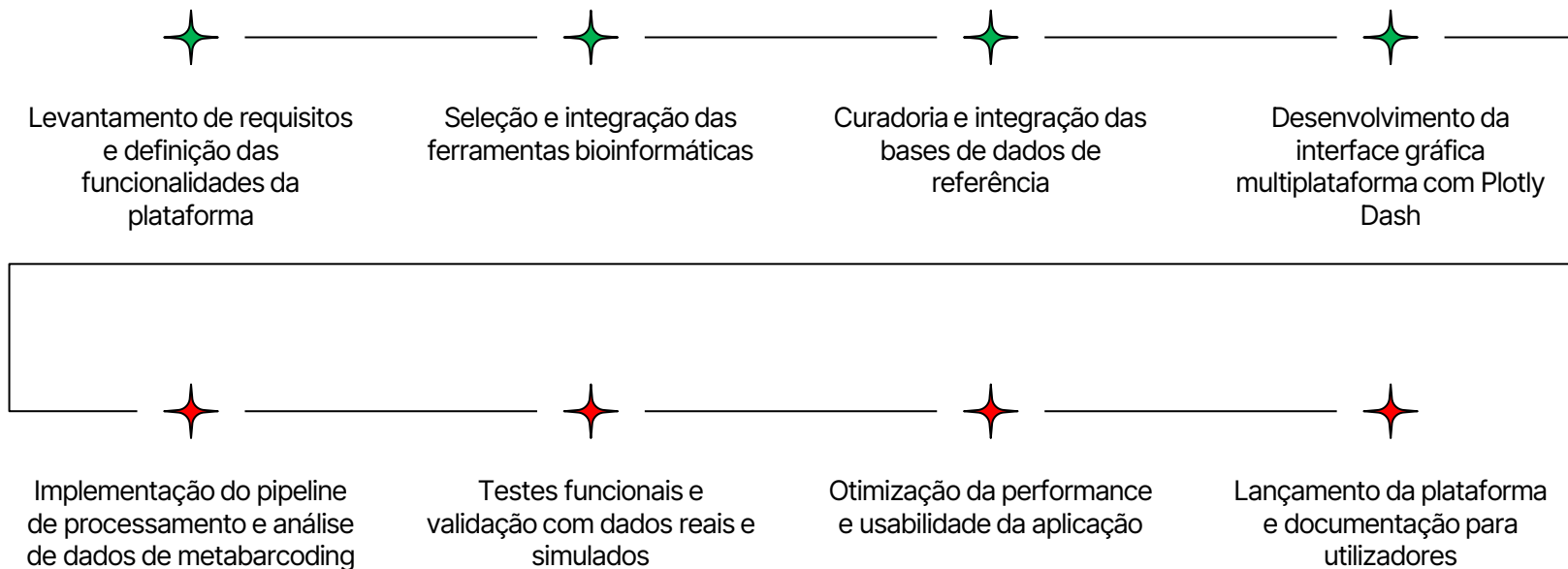
# Diagrama do workflow da plataforma de metabarcoding

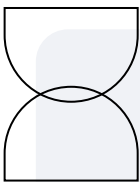


**Fig.5-** Diagrama que demonstra o workflow da plataforma de metabarcoding.



# Etapas do Desenvolvimento da Plataforma de Metabarcoding





Dash

127.0.0.1:8050

# Metabarcoding Platform for Fish Parasite Detection

1. Upload Data

2. Quality Control

3. Preprocessing

4. Sequence Analysis

5. Reference Databases

6. Phylogenetic Analysis

7. Results

Upload Raw Sequencing Data


Drag and Drop or Click to Select .fastq.gz Files





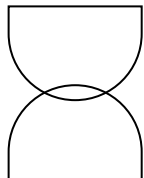
05

# Resultados Esperados

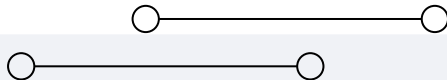




- Desenvolvimento de uma solução de **software de metabarcoding intuitiva, multiplataforma** e melhorada em relação ao PMiFish.
- **Maior precisão na classificação de espécies** e simplificação de procedimentos bioinformáticos complexos.
- Integração de uma interface fácil de usar com uma **biblioteca de referência de sequências de parasitas de peixes** cuidadosamente selecionada.
- Melhoria na **monitorização e controlo de comunidades de parasitas de peixes**.
- Proporcionar **benefícios para:**
  - Investigadores.
  - Reguladores de segurança alimentar.
  - Indústria da aquacultura.







# Obrigado



Universidade do Minho  
Escola de Engenharia



cbma  
centro de **biologia**  
molecular e ambiental



João Faria

Departamento de  
Informática, Universidade  
do Minho



João Carneiro

Centro de Biologia Molecular e Ambiental (CBMA), Departamento de  
Biologia, Universidade do Minho

Instituto de Ciência e Inovação para a Bio-Sustentabilidade (IB-S),  
Universidade do Minho



Teresa Rito

