

PARAFISH: A Cross-Platform Graphical Interface for Reproducible Metabarcoding Analysis in Aquatic Parasite Monitoring



Universidade do Minho
Escola de Engenharia



JOÃO FARIA, MSC STUDENT IN BIOINFORMATICS, SCHOOL OF ENGINEERING, UNIVERSITY OF MINHO
TERESA RITO (CBMA/DEPT. BIOLOGY, UMINHO; IB-S, UMINHO)
JOÃO CARNEIRO (UNIVERSIDADE CATÓLICA PORTUGUESA-CBQF-ESB)

BACKGROUND

Aquatic parasites are ecologically relevant indicators of ecosystem health and fish population dynamics. Although DNA metabarcoding enables high-throughput detection, many existing pipelines require advanced command-line skills, fragmented tools, and manual data handling. These barriers limit reproducibility, comparability across studies, and adoption by non-specialist users.

OBJECTIVE

We developed PARAFISH, a user-friendly, graphical platform that integrates quality control, preprocessing, sequence analysis, taxonomic assignment, and phylogenetic reconstruction into a single reproducible workflow.

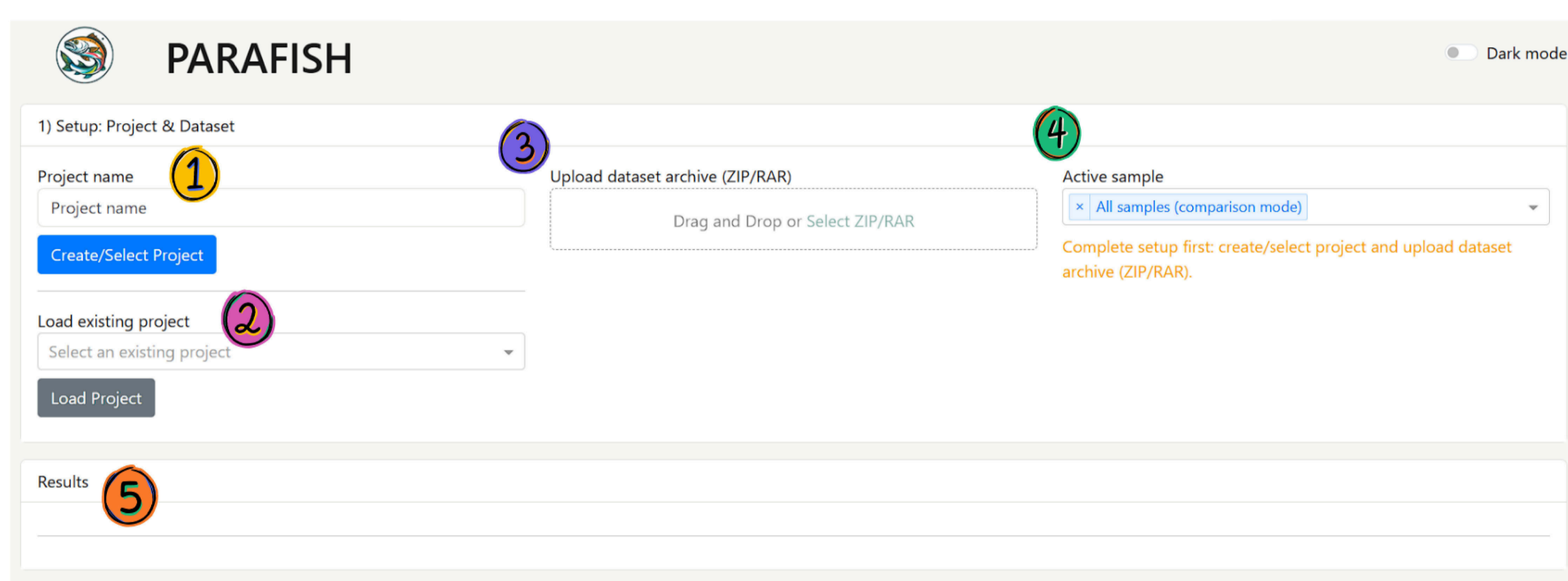


Figure 1. PARAFISH graphical interface.

(1)(2) Project initialization:

Create or load a project to preserve run history and settings.

(3) Dataset ingestion:

Upload raw sequencing files in a single archive for automatic indexing.

(4) Sample scope selection:

Choose active sample(s) or compare all samples in multi-sample mode.

(5) Integrated results workspace:

All outputs are generated in one place with reproducible project organization.

KEY CONTRIBUTIONS

- End-to-end workflow integration from FASTQ input to publication-ready outputs.
- Multi-sample project and run management with session persistence.
- Flexible taxonomy support with curated and custom reference databases.
- Built-in generation of community matrices and comparative summaries.
- Exportable outputs for downstream ecological and statistical analyses.

METHODOLOGY

The platform is implemented in Python using a Dash-based web interface. It orchestrates established bioinformatics tools for QC, trimming/merging, OTU/ASV processing, taxonomic classification, and tree inference. Database workflows include curated references and custom user-provided datasets, enabling adaptable use across study designs.

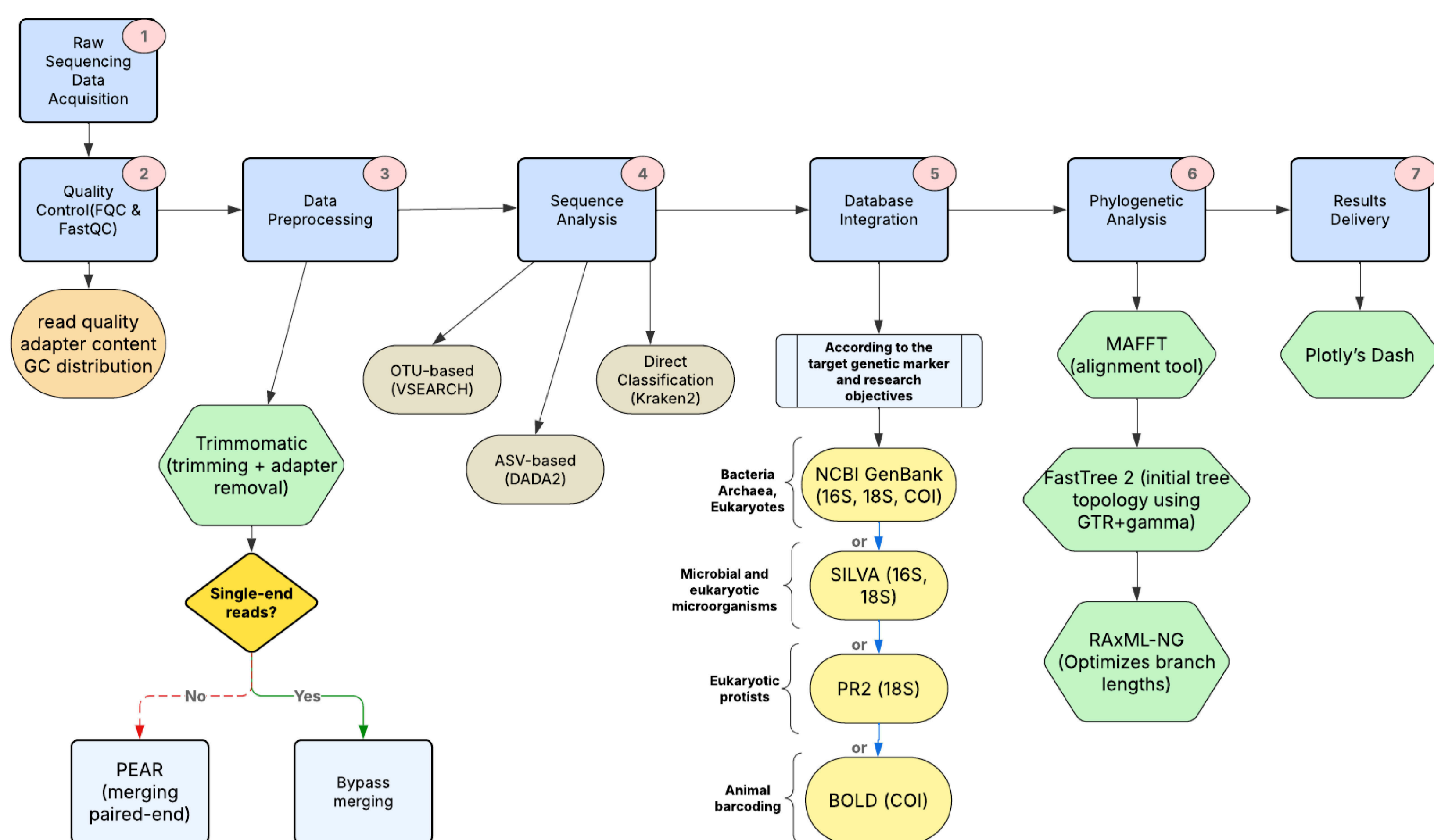


Figure 2. Workflow diagram of the metabarcoding platform.

RESULTS AND OUTPUT REPRESENTATION

PARAFISH produces standardized outputs including:

- (1) QC summaries,
 - (2) filtered/processed sequence files,
 - (3) taxonomic assignment tables and best-hit summaries,
 - (4) community matrices (presence/absence and abundance),
 - (5) phylogenetic trees and alignment files,
 - (6) downloadable session exports for reproducible reporting.
- In multi-sample mode, these outputs enable direct cross-sample comparisons and support ecological interpretation.

Table 1. Representative community matrix (Presence/Absence) for multi-sample comparison.

	SAMPLE_1	SAMPLE_2	SAMPLE_3	SAMPLE_4	SAMPLE_5
ANISAKIS SP.	1	1	0	1	0
CONTRACAECUM SP.	1	0	1	1	1
HYSTEROETHYLACIUM SP.	0	1	1	0	1
PSEUDOTERRANOVA SP.	1	1	1	0	0
ACANTHOCEPHALA SP.	0	0	1	1	1

Rows represent detected parasite taxa and columns represent individual samples; values indicate presence (1) or absence (0). The data displayed are examples only and do not reflect actual study results.

PARAFISH lowers the technical barrier to robust metabarcoding analysis while preserving methodological rigor. By combining guided interaction, reproducible project organization, and interoperable outputs, the platform supports both specialist and non-specialist users in aquatic parasite monitoring workflows.

Future development will focus on benchmarking across larger datasets, expanding reference database support, and adding integrated statistical modules for diversity and differential community analyses.

FUNDING:

THIS WORK IS CARRIED OUT UNDER CONTRACT REF. 2023.15056.TENURE.043, FUNDED BY NATIONAL FUNDS THROUGH FCT - FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA, I.P. THIS WORK IS ALSO FUNDED BY NATIONAL FUNDS THROUGH FCT - FUNDAÇÃO PARA A CIÊNCIA E A TECNOLOGIA, WITHIN THE SCOPE OF UID/50016/2025 AND LA/P/0076/2020 ([HTTPS://DOI.ORG/10.54499/LA/P/0076/2020](https://doi.org/10.54499/LA/P/0076/2020)).