DEVELOPMENT OF ANANALYSIS PIPELINE
FOR THE IDENTIFICATION OF PROTEOMIC
BIOMARKERS OF THE CHEMICAL ' DEFENSOME' IN WILD FISH SPECIES.

Jordan DUTEL

Tidjani CISSE

Ines AMDOUNI

Master Bio-Informatic 2024-2025

CONTENTS

INTRODUCTION

❏ General Context

Freshwater ecosystems are under increasing environmental pressures, such as:

•**Chemical pollution** from industrial, agricultural, and domestic discharges,

•**Climate change**, which disrupts water's physical conditions,

•Exposure to **endocrine disruptors** and other toxic compounds.

These environmental stresses have a significant impact on fish health, making it essential to develop **biological tools** for monitoring and assessing their condition.

It is crucial to establish approaches that generate **reporter peptides** as **biomarkers** to evaluate responses to environmental stress.

Reporter Peptide 2

Alpha Helix

Reporter Peptide 1   Reporter Peptide 3
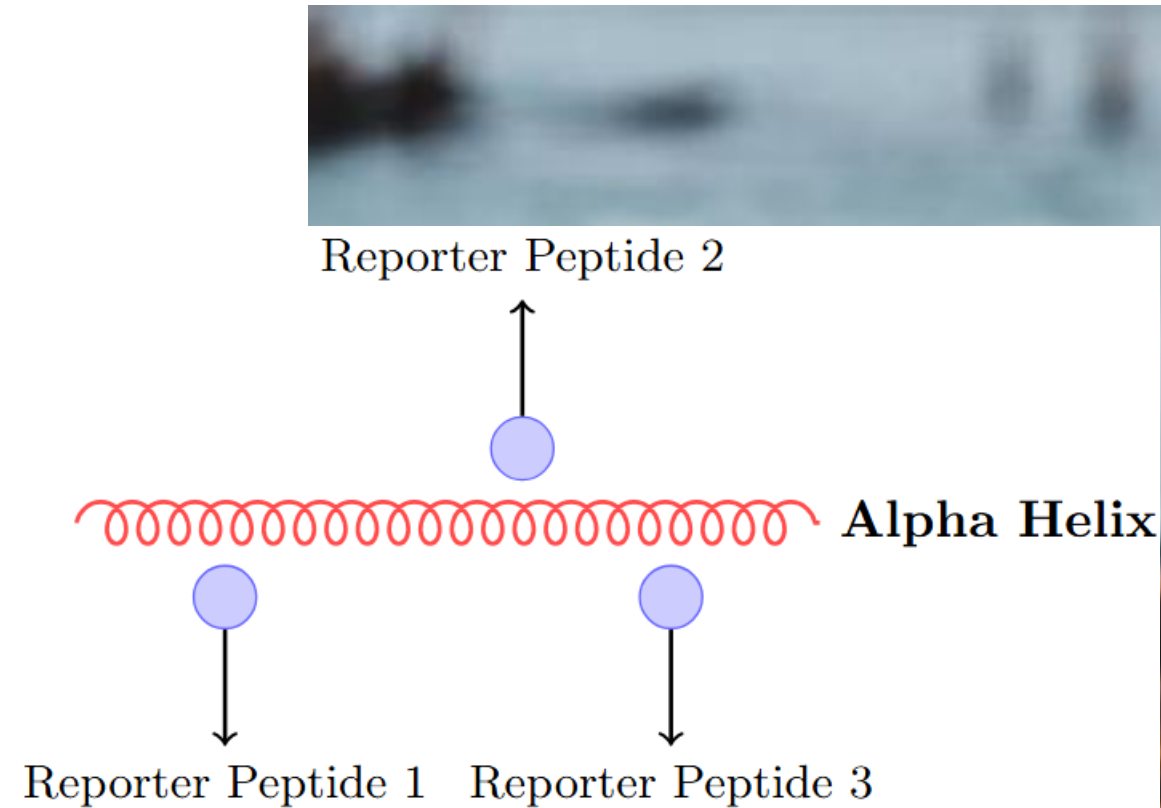
Reporter peptide on the alpha helix

INTRODUCTION

❑ Problematic

"How to identify stress biomarkers in local freshwater species"

INTRODUCTION

❑ Objectives

This project aims to:

• Identify proteins and peptides of interest involved in **stress response mechanisms** (defensome) in freshwater fish.

• **In silico digestion** to prepare peptides for **mass spectrometry (MS)**.

• Build an **automated bioinformatics pipeline.**

• Develop an **interactive application (R Shiny)** to easily explore and visualize the results.

DATA

Gobio gobio
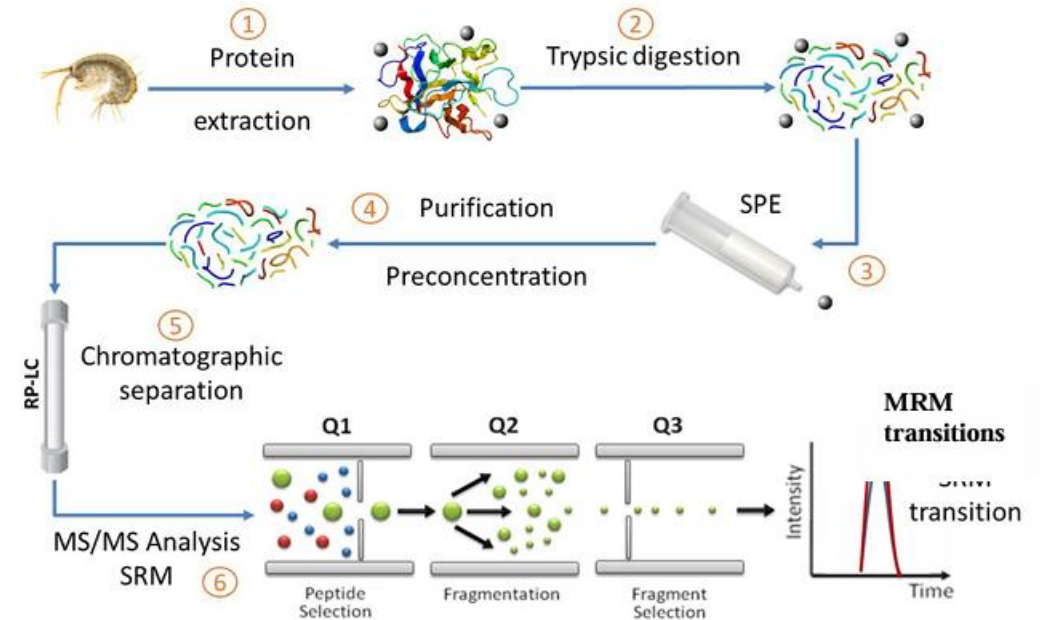
Barbus barbus

❑ Data sources :

•Public databases: UniProt, NCBI

•File formats: FASTA (protein and genome sequences)

Oncorhynchus mykiss

Squalius cephalus
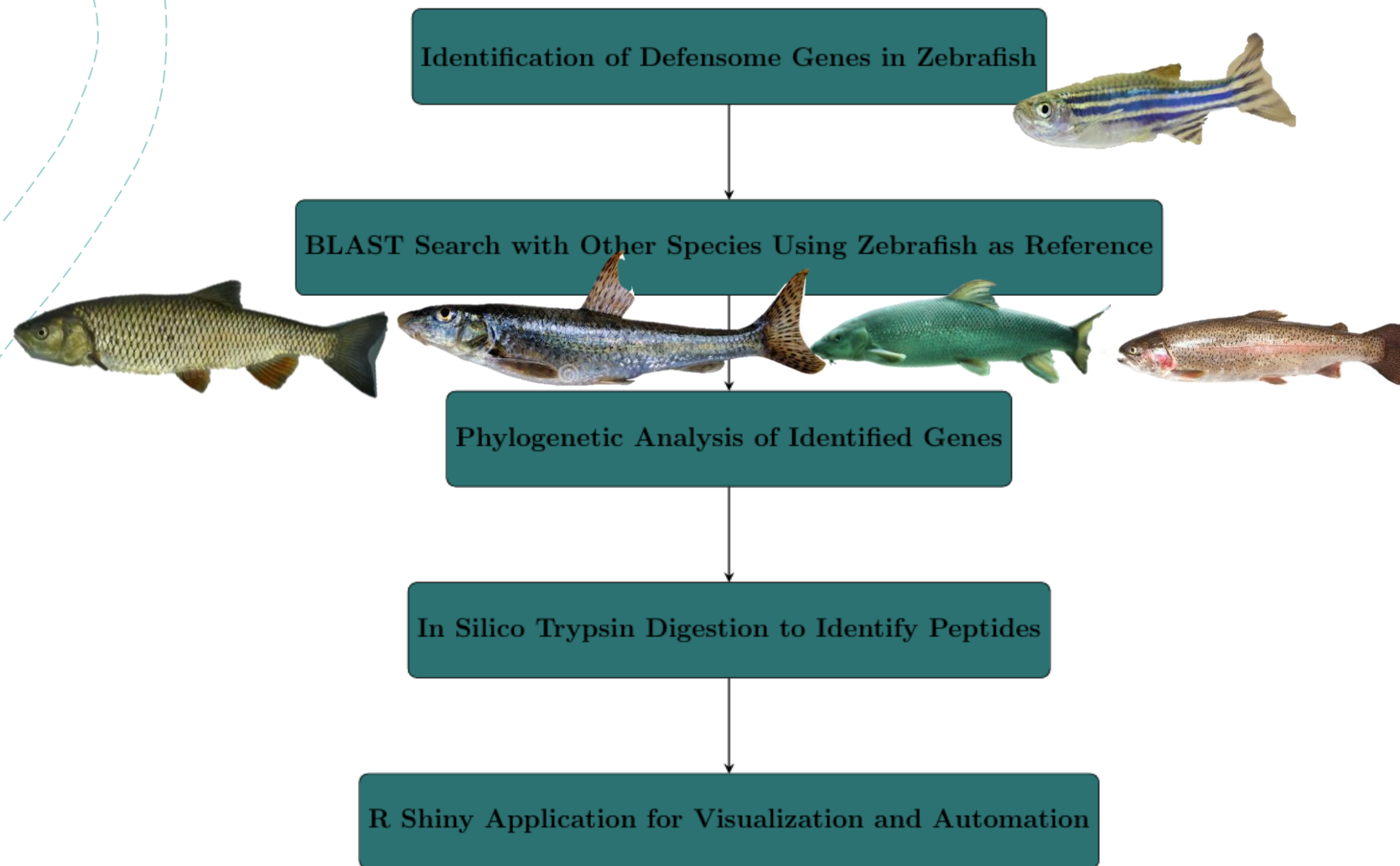

Danio rerio

WORKLOW

Identification of Defensome Genes in Zebrafish

BLAST Search with Other Species Using Zebrafish as Reference

Phylogenetic Analysis of Identified Genes

In Silico Trypsin Digestion to Identify Peptides

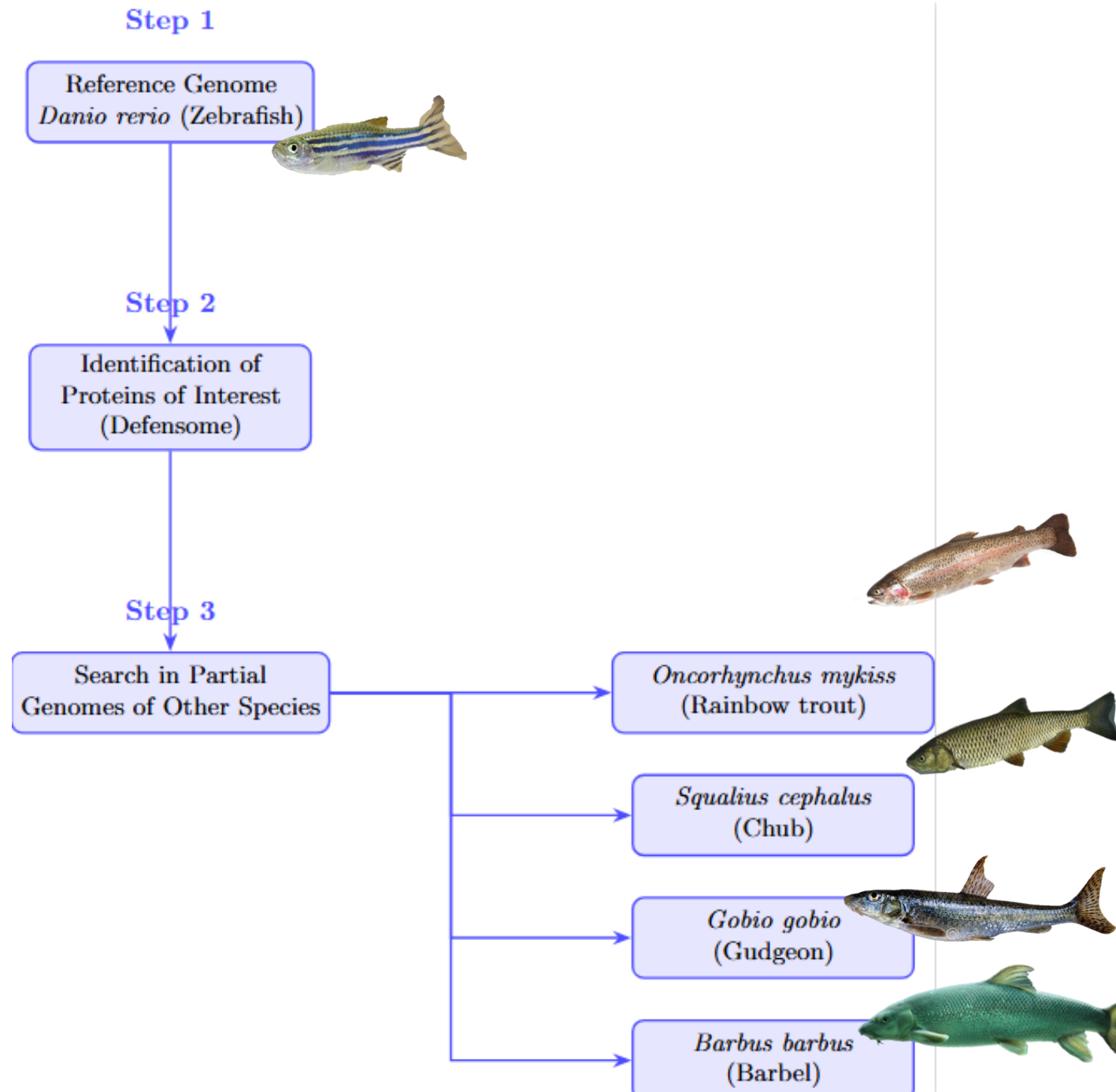R Shiny Application for Visualization and Automation

STEP 1

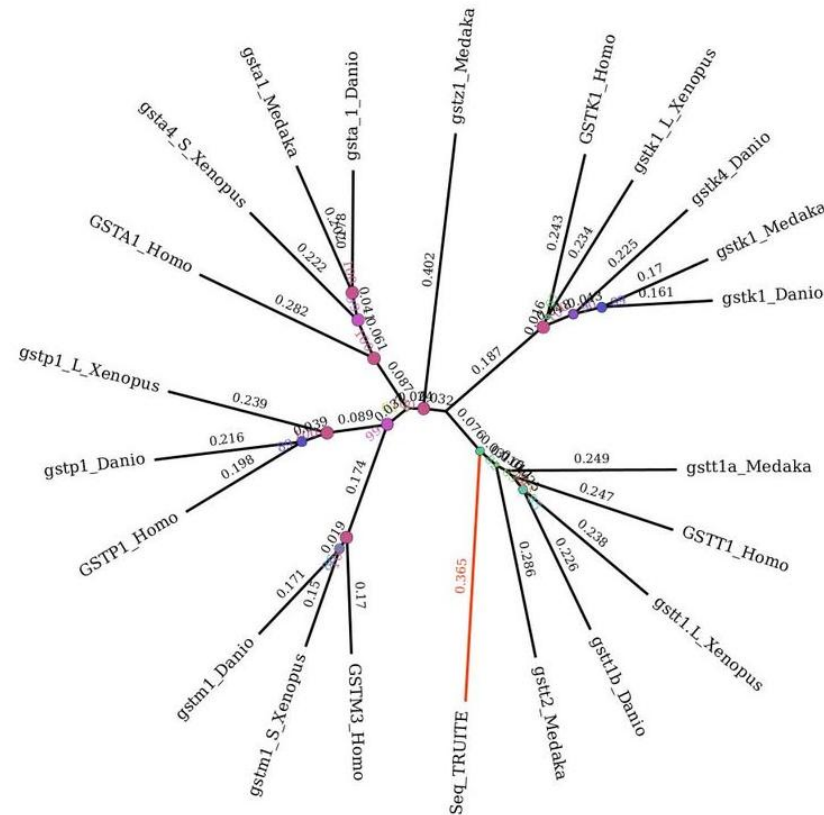- Genome alignment and identification of proteins of interest

- <u>BLAST</u>

STEP 2

- Phylogenetic analysis of identified proteins

- Phylogenetic analysis highlights the evolutionary relationships and conservation of defensome proteins.

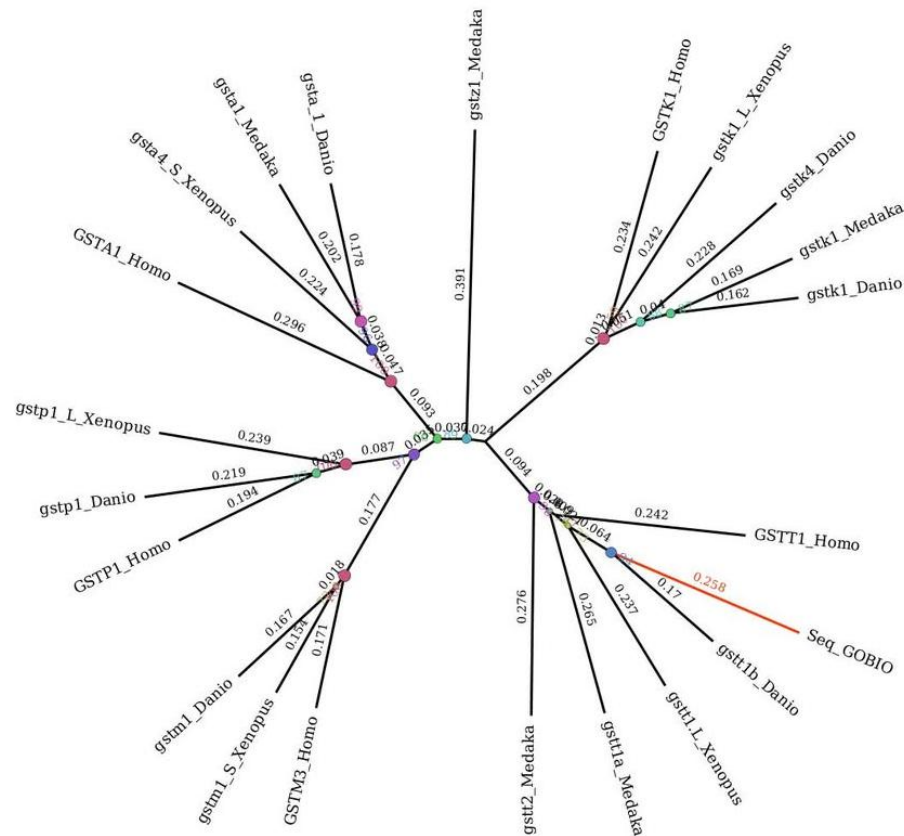|	alignments
1	>gsta_1_Danio
2	------MSG-KVVLHYFNGR-GKMESIRWLLAAAGVQFEEVFLTEKEQ-------FD--K
3	L------------------LSDGALTFQQVPLVEIDGMK----LVQSKAILNYIAGKY
4	N----LYGKDLKERAMIDIYSEGLID-LMEM----IMVSPF-----TPAENKE-KVFSNI
5	EEKAKVRFLPVF-EKALA----NSSFLVGKQLSRADVHLLEATLMLQELFPSILATFPKI

GST *Oncorhynchus mykiss*

STEP 2

- Phylogenetic analysis of identified proteins

- Phylogenetic analysis highlights the evolutionary relationships and conservation of defensome proteins.

alignments

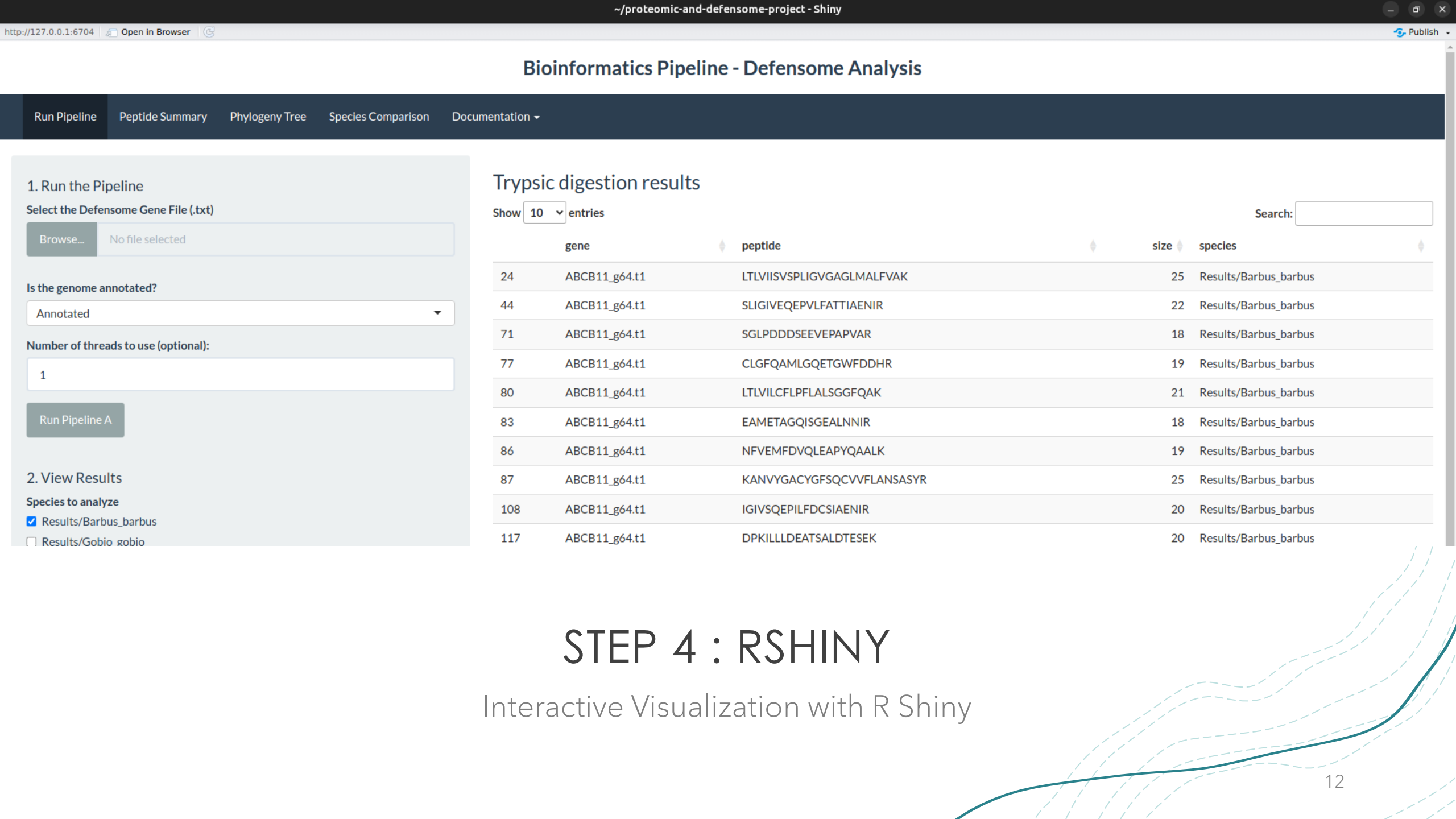| 1 | >gsta_1_Danio |
2	------MSG-KVVLHYFNGR-GKMESIRWLLAAAGVQFEEVFLTEKEQ-------FD--K
3	L-----------------LSDGALTFQQVPLVEIDGMK----LVQSKAILNYIAGKY
4	N----LYGKDLKERAMIDIYSEGLID-LMEM----IMVSPF-----TPAENKE-KVFSNI
5	EEKAKVRFLPVF-EKALA----NSSFLVGKQLSRADVHLLEATLMLQELFPSILATFPKI

GST *Gobio Gobio*

STEP 3

- Identification of peptides of interest.

- In *silico* trypsin digestion to obtain a list of peptides of interest

ID	Peptide Sequence	Length (aa)
24	LTLIVISVPLIGVGAGLMALFVAK	25
44	SLIGVEQPLVFATTIAENIR	22
71	SGLPDDDSEEVEAPPVAR	18
77	CLGFQAMLGQETGWFDDHR	19
80	LTLVLICFLPLALSGFQAK	21
83	EAMETAGQISGEALNNIR	18
86	NFVEFMDQVLEAPYQAALK	19
87	KANVYGACFGSQCVFLANASSYR	25
108	IGIVSQEPILFDCSIAENIR	20
117	DPKILLDEATSALDTSEK	20

STEP 4 : RSHINY

Interactive Visualization with R Shiny

STEP 4 : RSHINY

Interactive Visualization with R Shiny

Peptide Length Distribution

Peptide Length Distribution
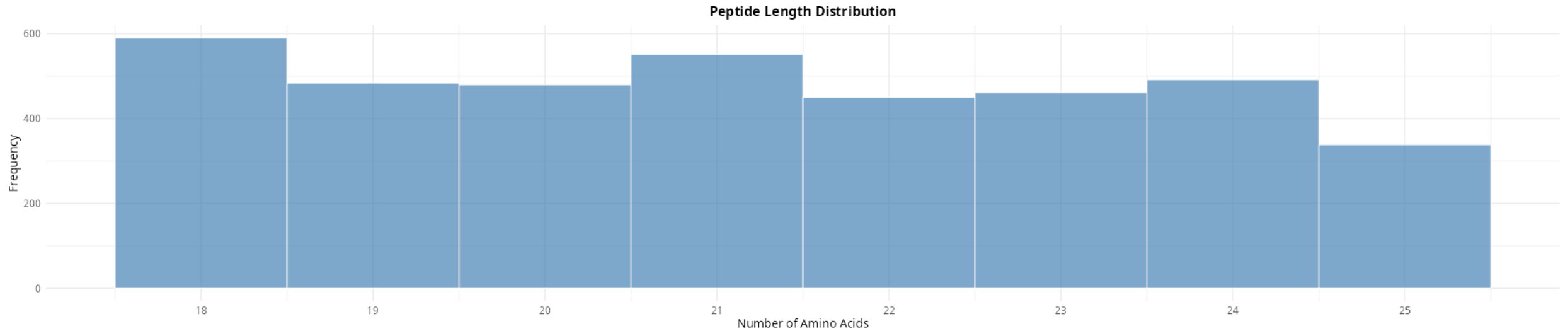
Statistical Summary

gene	Total_Peptides	Min_Length	Max_Length	Mean_Length	Median_Length	SD_Length
ABCB11_g64.t1	10	18	25	20.70000	20.0	2.5841397
ABCB6_g8.t1	7	18	23	19.57143	19.0	1.7182494
ABCC8_g43.t1	13	18	24	21.46154	21.0	1.5607362

STEP 4 : RSHINY

Interactive Visualization with R Shiny
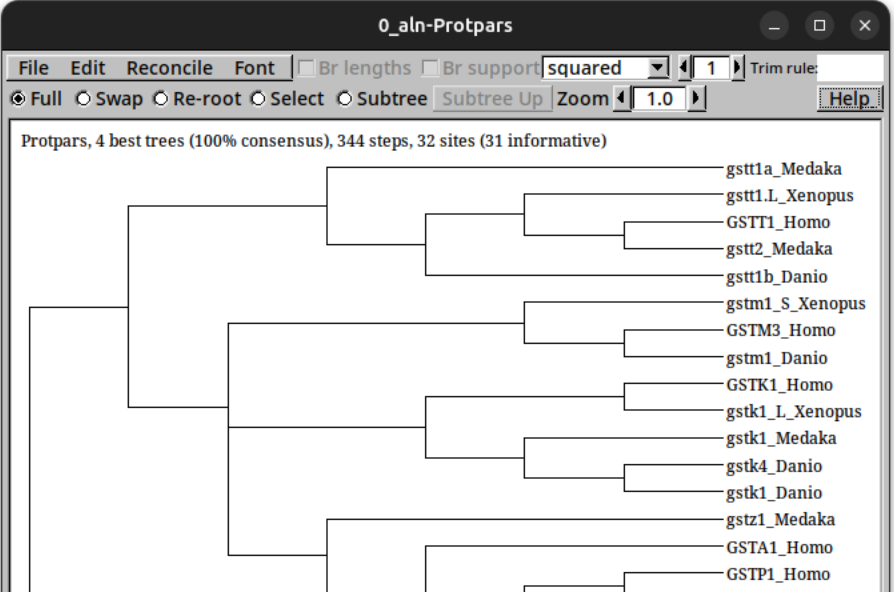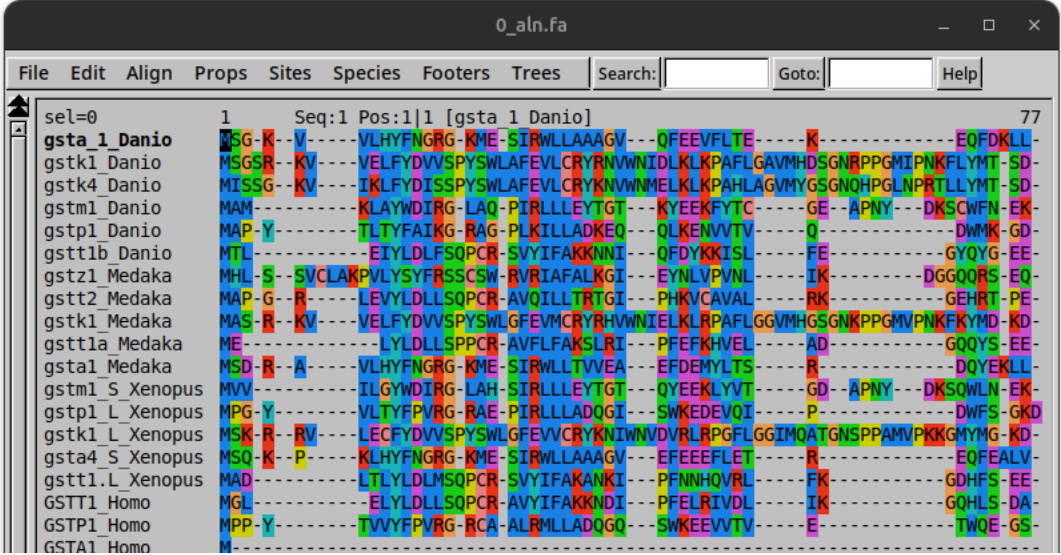
STEP 4 : RSHINY

Interactive Visualization with R Shiny

This tab helps users understand the output generated by the application.

1. **Peptide Results:** Explanation of each column in the peptide table.
2. **Filtering Options:**
   - **Gene-Specific Peptides:** Filters peptides uniquely aligned to a single gene.
   - **Species-Specific Peptides:** Filters peptides uniquely aligned to the selected species.
3. **Visualization:** Graphical representations, such as bar charts or heatmaps.
4. **Export Options:** Guidance on exporting results in common formats for downstream analyses.

For more information, please refer to the README.md file of the project.

STEP 4 : RSHINY

Interactive Visualization with R Shiny

CONCLUSION

Pipeline extension to other species to assess the health of broader freshwater ecosystems.

- We have designed and implemented an automated bioinformatics pipeline to generate a list of peptides of interest.
- To validate this pipeline, we focused on four local freshwater species:
- *Oncorhynchus mykiss* (rainbow trout),
- *Squalius cephalus* (chub),
- *Gobio gobio* (gudgeon),
- *Barbus barbus* (barbel).
- This pipeline is designed to be adaptable to other species, offering great flexibility.

FUTUR STEP

The peptide list generated by the pipeline can be used for targeted analyses in **mass spectrometry**. A workflow using tools like Skyline will allow to:

Quantify the peptides of interest

Validate their potential role as **stress biomarkers**.