Problem definition

- Data in bioinformatics and tools that are used to manipulate that data are being generated at a very fast rate.
- In scientific research, one of the fundamental principles is the independent verification of data through the re-creation of experiments in order to arrive at the same conclusions. However, there is a lack of effort made in scientific research to make results reproducible which leads to a difficulty to reuse the data.
- The FAIR guiding principles provide guidelines to ensure that data can be found and used by machines, in turn supporting data reuse by individuals. This is done through ensuring that data follows the four principles provided by FAIR:
  - Findability,
  - Accessibility,
  - Interoperability, and
  - Reusability.

What has been achieved so far:

- The main goal was to create a tool that semi-automates the FAIR assessment of bioinformatics tools and datasets, and allowing users to refine the score through a portal.
- This was achieved through the use of web crawling techniques to obtain information that satisfies metrics provided by the FAIRshake tool rubric and the FAIRshake dataset rubric. A score was then calculated based on which metrics were satisfied.
- The user can then see summary information of the tool along with a FAIR score and what can be done to improve that score. The portal also allows users to define pipelines and calculate their FAIR scores based on the tools and datasets used.

Screenshots of the system:

| | Tool Name | Findability | Accessibility | Interoperability | Reusability |
|---|---|---|---|---|---|
| | BLAST | 90.0% | 100.0% | 50.0% | 50.0% |
| | FASTA | 100.0% | 50.0% | 100.0% | 100.0% |
| | Clustal Omega | 90.0% | 100.0% | 100.0% | 100.0% |
| | HAMMOCK | 100.0% | 100.0% | 50.0% | 75.0% |
| | PyroHMMvar | 100.0% | 100.0% | 50.0% | 100.0% |
| | CytoSPADE | 100.0% | 0.0% | 50.0% | 75.0% |
| | MEANS | 100.0% | 100.0% | 100.0% | 75.0% |

## Summary information of tool

**BLAST**

| Findability | | 90.0% |
|---|---|---|
| Download Link | Link provided on website. | 40.0% |
| DOI | https://doi.org/10.1002/cpet.8 | 25.0% |
| Description | The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. | 25.0% |
| Older Versions | Previous versions of tool unavailable. Findability could be improved if previous versions of tool were made available. | 0.0% |

Refine Findability

| Accessibility | | 100.0% |
|---|---|---|
| API availability | Tool supports API usage. | 50.0% |
| Command Line | Tool can be accessed through command line. | 50.0% |

Refine Accessibilitiy

## User Refinement

## BLAST - Findability

Download Link
`https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=E`

DOI Link
`https://doi.org/10.1002/cpet.8`

Description
The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

Previous Versions
○ Previous Versions available on website/public repository.
◉ Previous Versions unavailable.

Update

## Pipeline FAIR score

| Pipeline Name | Findability | Accessibility | Interoperability | Reusability |
|---|---|---|---|---|
| Conflicting Evolutionary Patterns Due to Mitochondrial Introgression and Multilocus Phylogeography of the Patagonian Freshwater Crab Aegla neuquensis | 93.12% | 68.75% | 87.5% | 76.56% |