Building EMERALD-UI

a protein alignment visualization tool for EMERALD

Presenter: Andrei Preoteasa
Research Group: Algorithmic Bioinformatics Lab
Project: emerald-ui

Presentation Structure

What We'll Cover Today

1. **Background & Motivation**
2. **Limitations of Traditional Methods**
3. **EMERALD Algorithm**
4. **EMERALD UI Overview**
5. **Technical Implementation**
6. **EMERALD-UI In Action**

Background & Motivation

Why Protein Alignment Matters

Proteins are the workhorses of biology, and comparing their sequences helps us understand function, evolution, and disease.

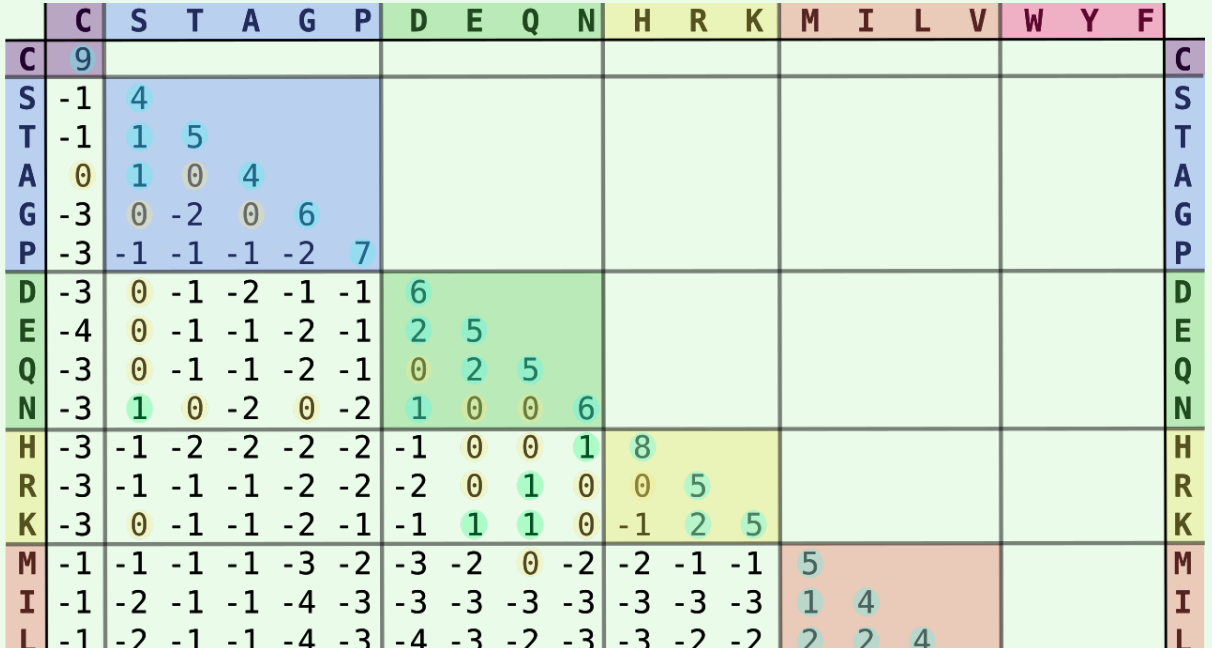
Example Protein Alignment:

```
A: MKTAYIAKQRQISFVKSHFSRQ
B: MKTA--AKQRQISFVKSHFSRQ

Match: +5   Mismatch: -2   Gap: -6
```

How Are Proteins Aligned?

- **Sequence alignment algorithms** compare amino acid sequences to find similarities.
- **Cost matrices** (e.g., BLOSUM, PAM) assign scores for matches, mismatches, and gaps.
- The goal: **maximize total score** by aligning similar regions and penalizing differences.



	C	S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	Y	F	
C	9																				C
S	-1	4																			S
T	-1	1	5																		T
A	0	1	0	4																	A
G	-3	0	-2	0	6																G
P	-3	-1	-1	-1	-2	7															P
D	-3	0	-1	-2	-1	-1	6														D
E	-4	0	-1	-1	-2	-1	2	5													E
Q	-3	0	-1	-1	-2	-1	0	2	5												Q
N	-3	1	0	-2	0	-2	1	0	0	6											N
H	-3	-1	-2	-2	-2	-2	-1	0	0	1	8										H
R	-3	-1	-1	-1	-2	-2	-2	0	1	0	0	5									R
K	-3	0	-1	-1	-2	-1	-1	1	1	0	-1	2	5								K
M	-1	-1	-1	-1	-3	-2	-3	-2	0	-2	-2	-1	-1	5							M
I	-1	-2	-1	-1	-4	-3	-3	-3	-3	-3	-3	-3	-3	1	4						I
L	-1	-2	-1	-1	-4	-3	-4	-3	-2	-3	-3	-2	-2	2	2	4					L
V	-1	-2	-1	0	-3	-2	-3	-2	-2	-3	-3	-2	-2	1	3	1	4				V
W	-2	-3	-2	-3	-2	-4	-4	-3	-2	-4	-3	2	-3	-1	-3	-2	-3	11			W
Y	-2	-2	-2	-2	-3	-3	-3	-2	-1	-2	2	-1	-2	-1	-1	-1	-1	2	7		Y
F	-2	-2	-2	-2	-3	-4	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	1	3	6	F
	C	S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	Y	F	

Issues in Protein Alignment

- **Gap Placement:** Deciding where to insert gaps affects biological interpretation.
- **Scoring Bias:** Choice of cost matrix influences which regions align.
- **Multiple Valid Alignments:** Different alignments may be equally plausible.
- **Functional Relevance:** Not all aligned regions are biologically meaningful.

Example Scenario:

```
Traditional:              ACDEFG--HIJK
Alternative (Same cost):  ACDE--FGHIJK
```

Instead of finding **one "optimal" alignment**, what if we could:

- **Explore the alignment landscape** - Find all good alignments, not just the best
- **Identify trustworthy regions** - Discover where alignments consistently agree
- **Reduce scoring bias** - See robust patterns across different parameters
- **Focus on functional relevance** - Highlight biologically meaningful regions

The EMERALD Approach:

> **If multiple high-scoring alignments agree on a region, that region is probably functionally important!**

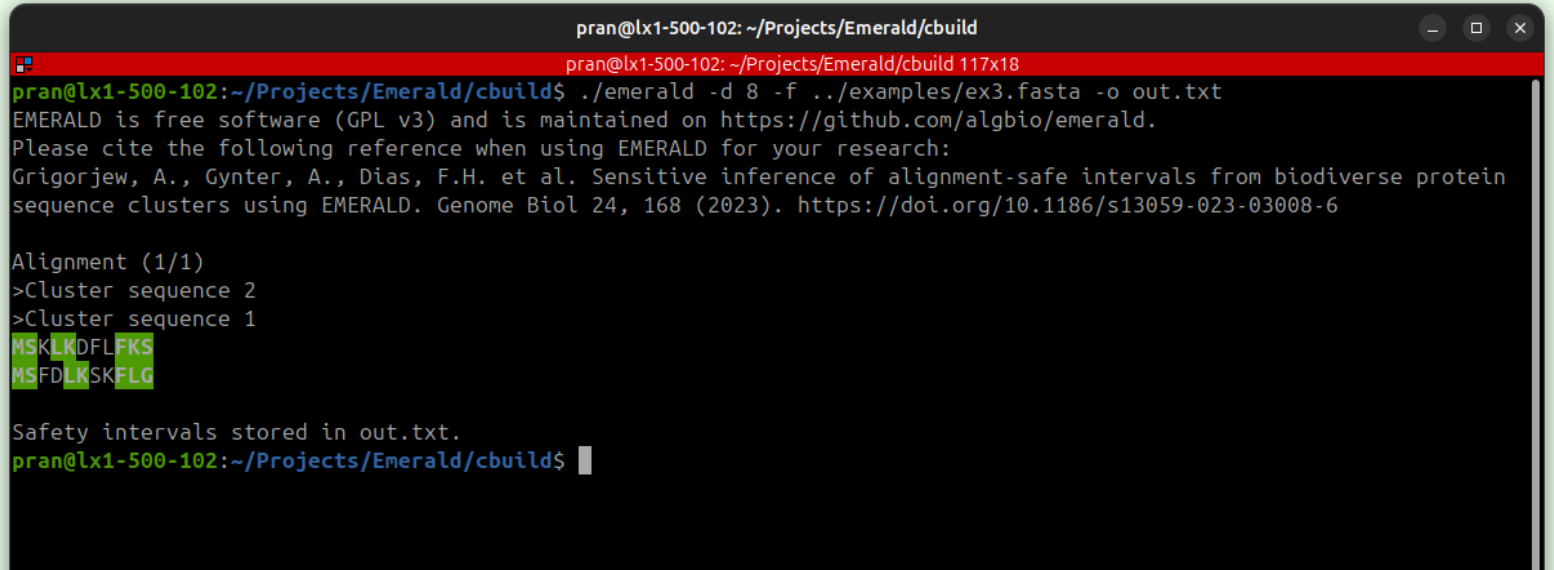
This is exactly what **EMERALD** does - it reveals which parts of your alignment you can trust.

EMERALD by itself

EMERALD is a terminal application for generating protein alingments written in C++.

Issues

- App only gives raw data, which is hard to interpret
- Only work in the terminal, which is hard to use
- Limited input methods



How EMERALD-UI Solves These Problems

- **Intuitive Visualization:** Draws an interactive, easy-to-use graph from raw EMERALD output.
- **Accessible Interface:** Runs in the browser, no terminal required.
- **Flexible Input:** Paste sequences, upload files, or search directly from UniProt.
- **Alignment Confidence:** Highlights common areas of multiple alignments.
- **Parameter Exploration:** Instantly adjust scoring and see how alignments change.
- **Safety Windows Visualization:** Highlights common regions in 3D structures of proteins.

Technology Stack

React Modern UI framework for building interactive web applications.	**TypeScript** Type-safe code for reliability and maintainability.	**WebAssembly (WASM)** Runs EMERALD executable in-browser.
D3.js Interactive, data-driven visualizations.	**UniProt API** Direct protein sequence search.	**Mol*** 3D protein structure visualization.

Demo Time

Let's see EMERALD-UI in action!

Thank You!

Questions?

Feel free to ask about EMERALD-UI, protein alignment, or anything from today's presentation.