JavaScript applications for bioinformatics — a project for a student helper

In the Section for Bioinformatics, Department of Health Technology, we need a student assistant with experience in JavaScript and HTML to help us make JavaScript-based versions of various web tools used in teaching. The project is to be carried out during the spring semester of 2023.

Background

When teaching the course 22111 Introduction to Bioinformatics, we use a long list of web servers to carry out various tasks. This works fine for a few participants, but with the number of students we have currently, server response time becomes a serious issue. This is hard to avoid in some cases which require access to large updated databases, but some of the computationally less heavy tasks could be carried out by the participants' own laptops instead of on backend servers, if they were implemented in JavaScript to be run in a web browser.

There is a web portal for bioinformatics named [Sequence Manipulation Suite](#) (SMS2), which is written 100% in HTML and JavaScript (GPL licensed). It contains a lot of tools for various lightweight DNA or protein sequence analysis tasks, but the individual tools are often not very flexible or only partially implemented. We would like better versions of a couple of the tools in SMS2, inspired by other tools. The user interface should be graphical, like that of SMS2.

Pairwise alignment

The SMS2 tools [Pairwise Align DNA](#) and [Pairwise Align Protein](#) are implementations of the [Needleman-Wunsch algorithm](#) for aligning two sequences globally by dynamic programming. However, they have a number of drawbacks:

1. they do not allow different values for gap opening and gap extension (affine gap penalties);
2. in the protein version, there are only a few substitution matrices available;
3. they contain no implementation of the [Smith-Waterman algorithm](#) for local alignment.

These missing features are all implemented in the command-line-based [EMBOSS](#) package (written in C, GPL licensed), in the programs Needle and Water. In the course, we use web server implementations of [Needle](#) and [Water](#) from the European Bioinformatics Institute (EBI), but when 180 students run these services simultaneously, waiting times become unacceptable.

The first subproject will be to use the source codes from SMS2 and EMBOSS as starting points for developing a JavaScript-based pairwise alignment browser interface.

DNA translation and ORF finding

VirtualRibosome is a web server, developed at DTU, for doing automated translation of DNA sequence to amino acid sequence. It also features an ORF (Open Reading Frame) finder which can find the longest stretch of codons that is not broken by a stop codon (this will often be the correct coding reading frame). The server works just fine, but when 100+ students use it at the same time, glitches in the communication between frontend and backend server can occur. Since the computations involved are quite lightweight, they could very well be carried out in the browser.

Some of the functionality of VirtualRibosome is found in the SMS2 tools Translate and ORF Finder, but we would prefer to have a more complete implementation that uses the output formats provided by VirtualRibosome.

The second subproject will therefore be to reimplement VirtualRibosome in JavaScript. The source code to the backend of VirtualRibosome, written in Python, will be provided. Possibly, some of the input options (GenBank and TAB format) could be left out, since they are not used very much, but that will depend on the available time.

Optional extra tasks

If there is still time left after implementing pairwise alignment and VirtualRibosome, we can discuss whether other services used in teaching, such as EasyPred or TreeHugger, can also be ported to JavaScript.

Links

- SMS2 implementation: https://www.bioinformatics.org/sms2/index.html
- SMS2 source code: https://github.com/paulstothard/sequence_manipulation_suite
- EMBOSS home page: https://emboss.sourceforge.net/ or http://emboss.open-bio.org/
- EMBOSS source code:
  ftp://emboss.open-bio.org/pub/EMBOSS/EMBOSS-6.6.0.tar.gz
- VirtualRibosome server:
  https://services.healthtech.dtu.dk/service.php?VirtualRibosome
- VirtualRibosome article:
  https://academic.oup.com/nar/article/34/suppl_2/W385/2505654

Requirements

- Experience with writing JavaScript and HTML.
- Some knowledge of C and Python.
- Experience with bioinformatics would be an advantage, but is not a requirement.