BioLovesData - Gene Analysis Program

Overview

BioLovesData is an integrated bioinformatics analysis tool using Tkinter GUI. It provides various gene and protein sequence analysis functions.

Main Features

1. Transversion Ratio Calculation

- Calculates the ratio of transitions and transversions in DNA sequences

- Input: DNA sequence files in FASTA format

- Output: Number of transitions/transversions and their ratio

2. Consensus Symbols Analysis

- Analyzes conservation in multiple sequence alignments

- Input: FASTA file or space-separated sequences

- Output: Conservation symbols for each position (* : identical, : : conservative, . : variable, space : gap)

3. Promoter Finder

- Detects bacterial promoter regions (-10, -35 boxes)

- Input: DNA sequences (FASTA or raw sequence)

- Options: Sequence length limit, maximum ambiguity setting

- Output: Location and sequence of discovered promoter regions

4. TM Domain Finder

- Predicts transmembrane domains

- Input: Protein sequences (FASTA or raw sequence)

- Options: Hydrophobicity threshold, hydrophobicity scale selection

- Output: Location, length, and hydrophobicity values of transmembrane domains

5. Multiple Sequence Alignment

- Performs multiple sequence alignment

- Input: Multiple sequences in FASTA format

- Options: Gap penalty setting

- Output: Aligned sequences and gap statistics

How to Use

1. Run the Program

Execute `BioLovesData.exe` (Onefile program. You can delete other files if you need only this one.)

Or

```Bash
python3 BioLovesData.py
```

2. File Selection

- Click "Select File" button in the top left to choose a file for analysis

- Supported formats: FASTA (.fasta, .fa), Text (.txt)

- Use "Preview File" button to check file contents

3. Analysis Function Selection

- Select the desired analysis function from the left panel

- Brief descriptions are provided for each function

4. Option Settings

- Related options are displayed based on the selected analysis

- Adjust parameters as needed

5. Run Analysis

- Click "Run Analysis" button to start the analysis

- Results are displayed in the right panel

6. Save Results

- Click "Save Results" button to save analysis results as a text file

Test Data

Test data provided with the program:

- `consensus test1.fasta` : for MSA conservation analysis testing

- `multi seq align test1.fasta:` for long MSA testing

- `multi seq align test2.fasta:` for simple MSA testing

- `multi seq align test3.fasta:` for aligned sequences testing

- `promoter test1 SUPER BIG... .fasta:` don't run it unless you have a fancy computer

- `promoter test2.fasta:` for finding promoter regions

- `TM domain test1.fasta:` for MSA transmembrain domains prediction testing

- `transversion ration test1.fasta:` for transition/transversion ratio calculation testing

System Requirements

- Python 3.11+

- tkinter (GUI library)

- numpy (numerical computation)

Installation

```bash
# Install tkinter (Ubuntu/Debian)
sudo apt-get install python3-tk

# Install numpy
pip3 install numpy
```

File Structure

```
BioLovesData/
├── BioLovesData.exe                   # exe file
├── BioLovesData.py                    # Main GUI program
├── calculate_transversion_ratio.py    # Transversion ratio analysis
├── consensus_symbols.py               # Consensus symbols analysis
├── find_promoter.py                   # Promoter finder
├── find_tm_domain.py                  # TM domain finder
├── multi_seq_alignment.py             # Multiple sequence alignment
├── Test Data                          # Contains test data
│   ├── consensus test1.fasta
│   └── ...
├── BioLovesData.pdf                   # User guide
└── BioLovesData.md                    # Same guide in markdown format
```

Notes

- Large files may take time to analyze

- FASTA file headers must start with '>'

- Protein sequences should use standard amino acid codes

Troubleshooting

- If the program doesn't run: Check tkinter installation

- If analysis errors occur: Check input file format

- If GUI doesn't display: Check DISPLAY environment variable setting

Version History

- v2.0: Interagated into one single program and added GUI

- v1.0: A collection of standalone Python scripts