Table 1: Revision History

Date	**Developer(s)**	**Change**
April 7th, 2025	Uriel Garcilazo Cruz	Initial version

User Guide for SubLiMat
Substitution Matrix Benchmarking Tool

Uriel Garcilazo Cruz

April 8, 2025

1 Introduction

SubLiMat is a bioinformatics tool for benchmarking substitution matrices using the Needleman-Wunsch global alignment algorithm. This guide provides instructions for installing and using the software.

2 System Requirements

- **Operating Systems**: Windows 10+, macOS 13+, or Linux
- **Python**: Version 3.8 or higher
- **Dependencies**: pandas, numpy

3 Installation

3.1 From Source

1. Clone the repository:

```
git clone https://github.com/UGarCil/UGarcil_capstone.git
```

2. Install dependencies:

```
pip install -r requirements.txt
```

4 Quick Start

4.1 Running the Program

Execute the main script from the command line:

```
from sublimat import file_manager
from sublimat.main import main

# To run the program:
# define the absolute path to:
# - substitution_matrices.txt
benchmark = "<path_to_your_file>/substitution_matrices.txt"
# - input_sequences.fasta
sequence_data = "<path_to_your_file>/input_sequences.fasta"
# - output directory
output_path = "<path_to_your_file>/data/"
# Execute the function composition export(main()) to export the
    results to the output directory
file_manager.export(main(sequence_data, benchmark), output_path)
```

Then execute the script using the Python interpreter. If your file is called main.py:

```
    python main.py
```

4.2 File locations

You need to specify the locations of your input files. The suggested locations are as follows:

- Input sequences: `data/input_sequences.fasta`
- Substitution matrices: `data/substitution_matrices.txt`
- Results output: `data/benchmark_results.csv`

5 Input File Formats

5.1 Sequence File (FASTA Format)

```
>Sequence1
ATGCGTACGT
>Sequence2
TGCGTACGTA
```

Requirements:

- Exactly two sequences
- Only characters A, T, C, G allowed
- Minimum length: 1 base pair

5.2 Substitution Matrix File

```
>Matrix1
1.0,-0.33,-0.33,-0.33
-0.33,1.0,-0.33,-0.33
-0.33,-0.33,1.0,-0.33
-0.33,-0.33,-0.33,1.0
>Matrix2
1.0,-1.0,-0.5,-1.0
-1.0,1.0,-1.0,-0.5
-0.5,-1.0,1.0,-1.0
-1.0,-0.5,-1.0,1.0
```

Requirements:

- Each matrix must be 4×4 (for A,T,G,C)
- Numeric values only
- One matrix per block (header + 4 lines)

6 Output

The program generates a CSV file with alignment results:

```
matrix,score
Matrix1,0.5
Matrix2,-12
...
```

Columns:

- **matrix**: Name of substitution matrix
- **score**: Alignment score (lower is better)

7 Troubleshooting

Table 2: Common Issues and Solutions

Issue	**Solution**
"Invalid FASTA File Error"	Ensure file has exactly two sequences with headers starting with `">"`
"Invalid Substitution Matrix"	Verify all matrices are 4×4 and contain only numbers
"Zero-length sequences"	Check that sequences contain at least one base pair

8 Examples

8.1 Basic Example

You can find a basic example of execution in the following demo: HERE.

9 Contact

For support or questions, contact:

- Uriel Garcilazo Cruz
- Email: garcilau@mcmaster.ca
- GitHub: `https://github.com/UGarCil`