Barcoding Processor User Manual

1. Overview

The Barcoding Processor is a comprehensive toolkit designed for the processing and analysis of DNA/RNA barcode segments. It facilitates various functionalities including batch sequence merging, degenerate base processing, specificity and recall rate preprocessing, recall rate processing, specificity processing, 1D barcode segment visualization, and primer design. This user manual provides detailed instructions on how to utilize the toolkit efficiently, aiming to assist researchers in the field of molecular biology and genetics.

2. Installation

Before running the Barcoding Processor, ensure you have Python installed on your system (version 3.6 or later is recommended). This application relies on Tkinter for the GUI, Pandas for data handling, and python-docx for document manipulation, which are included in standard Python installations or can be easily installed via pip.

3. User Interface

Upon launching the application, the user is presented with a main window divided into different sections, each corresponding to a specific functionality. The menu bar at the top allows users to navigate between functionalities.

3.1. Menu Bar

- Merge: Batch process sequences by merging multiple FASTA files.
- DeleteDBs: Process sequences by removing degenerate bases.
- SandPpreprocessing: Preprocess sequences for specificity and recall rate calculation.
- Recall rate: Calculate the recall rate for given barcode segments against species sequences.
- Specificity: Calculate the specificity for given barcode segments against species sequences.
- Barcodemaking: Visualize barcode segments as 1D codes.
- Primermaking: Design primers based on barcode segments.

3.2. File and Folder Browsing

- Browse File: Allows the user to select a single file (e.g., FASTA, TXT, DOCX) from their file system.
- Browse Folder: Enables the selection of a directory, facilitating operations on multiple files or storing output files.

3.3. Sequence Batch Merging (Merge)

- Input folder path: Select the folder containing FASTA files to be merged.
- Output folder path: Choose the destination folder for the merged sequence file.
- Click “Start Process” to begin merging.

3.4. Simplified Base Processing (DeleteDBs)

- Input: Requires a sequence file path in FASTA, FAS, or TXT format.

- Output: Specifies the path where the processed file, with simplified bases, will be saved.
- Click “Start Process” to remove degenerate bases from sequences.

3.5. Specificity and Recall Rate Preprocessing (SandPpreprocessing)
- Input: Accepts a path to a sequence file containing both the sequences and barcode fragments.
- Output Directory: Designates a folder path to save the processed sequences, preparing them for specificity and recall rate calculation.
- Click “Start Process” to preprocess sequences for specificity and recall rate calculation.

3.6. Recall Rate Processing (Recall rate)
- Barcode File: Input file containing barcode sequences.
- Species Sequence Folder: Directory with species sequence files.
- Output File: Location to save the calculated recall rates.
- Click “Start Process” to calculate recall rates.

3.7. Specificity Processing (Specificity)
Similar to recall rate processing, but focused on calculating the specificity of the barcodes against the provided species sequences.
- Click “Start Process” to calculate specificity.

3.8. Barcode Visualization (Barcodemaking)
- Input File: Specifies the path to a document containing nucleotide sequences for visualization.
- Output File: Path to save the visually formatted barcode sequence document.
- Click “Start Process” to visualize barcode segments as 1D codes.

3.9. Primer Design (Primermaking)
- Input File: Text file containing nucleotide sequences for which primers are to be designed.
- Output File: Excel file path to save the designed primers and their properties.
- Click “Start Process” to design primers based on input sequences.

4. Running the Application
Start the application by executing the script in a Python environment. The graphical user interface (GUI) allows easy navigation between different functionalities, enabling users to process data efficiently.

4.1. Select the desired operation from the menu.
4.2. Fill in the required fields for each function.
4.3. Click the corresponding "Browse..." buttons to select files or folders.
4.4. Hit the "Start Process" button to execute the operation.

5. Troubleshooting
- Ensure all input files are in the correct format as specified.
- Check the output paths to ensure they point to valid directories or file names.
- For large datasets, processing times may be extended; ensure the application is not interrupted

during these operations.

6. Support

For technical support or to report bugs, please contact the software development team at [gs2022@hnu.edu.cn].

7. Citation

If you use the DNA Barcode Processor in your research, please cite it as follows:

You C, Jiang S, Ding Y, Ye S, Zou X, Zhang H, Li Z, Chen F, Li Y, Ge X, Guo X. RNA barcode segments for SARS-CoV-2 identification from HCoVs and SARSr-CoV-2 lineages. Virol Sin. 2024 Feb;39(1):156-168. doi: 10.1016/j.virs.2024.01.006. Epub 2024 Jan 20. PMID: 38253258; PMCID: PMC10877444.

This manual is intended to assist users in the effective utilization of the DNA Barcode Processor for their research purposes. For further details or updates, please visit the official website or contact the support team.