

Manual for Chimera_gen Script

Introduction

This Python script simulates gene transfer events between two genomes provided in FASTA files. It replaces a specified percentage of a random segment of the recipient DNA sequence with a random segment of the same length from the donor DNA sequence. The resulting chimeric genome is saved to a new FASTA file.

To run this script, you need to have the following libraries installed: Biopython, SeqIO, Seq, SeqRecord, and argparse.

The script can be executed from the command line and requires specific arguments to function correctly. The command line to run the script is:

```
python chimera_gen.py --donor donor.fna --recipient recipient.fna --percentage 10 --out chimeric_genome.fasta
```

Therefore, to run the command we need

- `chimera_gen.py` that is the name of the python script,
- `--donor` is a command line argument that specifies the path to the donor FASTA file, our donor fast file is `donor.fna`.
- `--recipient` specifies the path to the recipient FASTA file, in our case `recipient.fna`.
- `--percentage` specifies the percentage of the donor genome to be transferred to the recipient genome, in our case of example 10%.
- `--out` specifies the name of the output file where the chimeric genome will be saved, in our case `chimeric_genome.fasta` in `.fasta` format.

In the script we defined two functions: the `replace_sequence_random()` function and the `main()` function.

The `replace_sequence_random()` needs 4 parameters as input:

- The donor DNA sequence (`donor_seq`)
- The recipient DNA sequence (`recipient_seq`), that is the genome into which we want to add the donor DNA.
- The percentage of the sequence to be replaced (`percentage`).
- Path for the output file (`output_path`).

The purpose of the function is to validate the provided percentage. Calculates the length of the segment to be replaced based on the given percentage and replaces the specified portion of the recipient sequence with the corresponding portion from the donor sequence starting from random position. Creates a chimeric genome by concatenating the replaced segment with the remaining part of the recipient sequence. Writes the resulting chimeric genome to the specified output file in FASTA format.

In the script we defined also a function called `main()`, that serves as the entry point for the program when it is executed from the command line. Its primary role is to handle command line argument parsing, read the input sequences from the provided FASTA files, and initiate the sequence replacement process by calling the `replace_sequence_random()` function with the parsed arguments.

Notes:

Ensure the FASTA files provided contain valid DNA sequences. The script will raise a `ValueError` if the percentage is not between 0 and 100. The output FASTA file will have a description indicating the percentage of the donor sequence that was used in the replacement. By following this manual, you should be able to understand and execute the script effectively for gene transfer tasks between two DNA sequences.