Creating a Frankenstein's Genome

Helena Pound[1], Eric Gann[1], Steven W Wilhelm[2]

[1]University of Tennessee, Knoxville; [2]The University of Tennessee, Knoxville

Jun 29, 2021

1 *Works for me* Share dx.doi.org/10.17504/protocols.io.bv2zn8f6

The Aquatic Microbial Ecology Research Group - AMERG (The Buchan, Zinser and Wilhelm labs)

Great Lakes Center for Fresh Waters and Human Health

Helena Pound
University of Tennessee, Knoxville

ABSTRACT

This short, command-line protocol is used to combine coding sequences (nucleic acids) from reference genomes into a single file with all coding sequences, with customizable clustering levels.

DOI

dx.doi.org/10.17504/protocols.io.bv2zn8f6

PROTOCOL CITATION

Helena Pound, Eric Gann, Steven W Wilhelm 2021. Creating a Frankenstein's Genome. **protocols.io**
https://dx.doi.org/10.17504/protocols.io.bv2zn8f6

CREATED

Jun 23, 2021

LAST MODIFIED

Jun 29, 2021

PROTOCOL INTEGER ID

51001

BEFORE STARTING

Please download CD-HIT before beginning this protocol.

1 Download the nucleic acid coding sequences from all reference genomes you wish to include in your Frankenstein's genome in a .fasta format.

2 Combine all downloaded fasta files.

Concatenate fasta files

```
cat *.fasta > output.fasta
```

Citation: Helena Pound, Eric Gann, Steven W Wilhelm (06/29/2021). Creating a Frankenstein's Genome. https://dx.doi.org/10.17504/protocols.io.bv2zn8f6

2.1 The linux cat command concatenates all files, in this case all files ending with .fasta in the working directory you are in. The output.fasta file will be the concatenated file to be used in the next step.

3

CD-HIT

```
cd-hit-est -i X1 -o X2 -c X3 -n X4
```

Used to combine reference genomes by threshold.

Linux

3.1 X1 is the concatenated reference .fasta file, X2 is the output folder name, X3 is the clustering threshold, and X4 is the word size.

3.2 Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150-3152.

Citation: Helena Pound, Eric Gann, Steven W Wilhelm (06/29/2021). Creating a Frankenstein's Genome. https://dx.doi.org/10.17504/protocols.io.bv2zn8f6