

Dec 09, 2024

Concatenate Sequenced Data for Bioinformatics Analysis

DOI

<https://dx.doi.org/10.17504/protocols.io.81wgbr6jnlpk/v1>

Dwi Cahyani¹

¹NCSU



Dwi Cahyani

NCSU

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.81wgbr6jnlpk/v1>

Protocol Citation: Dwi Cahyani 2024. Concatenate Sequenced Data for Bioinformatics Analysis. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.81wgbr6jnlpk/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: December 02, 2024

Last Modified: December 09, 2024

Protocol Integer ID: 113433

Keywords: concatenate sequenced data for bioinformatics analysis, concatenate sequenced data, sequenced data, bioinformatics, concatenated file, bioinformatics analysis, data, file, protocol

Abstract

This protocols helps to create concatenated files from a sequenced data.



Step 1: Prepare Your Data

- 1 **Go to the server** where your sequencing data is stored (e.g., Oxford Nanopore's EPI2ME or MinKNOW output directory)
- 2 **Download only the passed reads** (fastq.gz) because failed reads are typically of lower quality and are excluded from most analyses. Extracted all data.
- 3 **Make sure** all your FASTQ.gz files are stored in one folder for easy access during the concatenation process. Name the files systematically (e.g., sample1.fastq.gz, sample2.fastq.gz) to avoid confusion.

Step 2: Concatenate Using Galaxy

- 4 **Open the Galaxy Website**
Navigate to usegalaxy.org or the Galaxy server you're using.
- 5 **Log In or Create an Account**
If you don't already have an account, create one—it's free! Logging in will save your workflow and results.
- 6 **Upload Your Data** by clicking Upload data in the Galaxy interface. Drag and drop your FASTQ.gz files into the upload window or select them manually. Wait for the files to be fully uploaded and visible in your Galaxy history panel.
- 7 **Concatenate Datasets** by searching for "concatenate dataset" tool in the Galaxy tools panel. Use the "**Concatenate Dataset (tail to head cat)**" option for a fast and straightforward concatenation. Select your FASTQ.gz files in the order you want them concatenated, Run the tool!.
- 8 **Download the Concatenated File.** Once the process is complete, download the concatenated FASTQ.gz file to your local machine for further analysis.

Step 3: Concatenate Using the Windows Command Line

- 9 **Prepare Your Files**
Make sure all your FASTQ.gz files are in the same folder.
- 10 **Unzip the Files.**
If your FASTQ files are compressed (.gz), unzip them first. You can use a tool like **7-Zip** or run the following command in the terminal:
bash



Copy code

```
gzip -d *.gz
```

- 11 **Open Command Prompt.** On your Windows computer, press **Windows + R**, type cmd, and hit **Enter**
- 12 **Navigate to the Folder.** Use the cd command to move to the folder containing your files. For example:
cd C:\Users\YourName\SequencingData
- 13 **Concatenate the Files.** the command line write: type file1.fastq file2.fastq > concatenated.fastq
Replace file1.fastq and file2.fastq with the actual file names.
- 14
 - **Recompress the File.** If needed, recompress the concatenated file using:
gzip concatenated.fastq
- 15 Finish!