Organization	Course	Exercise	Semester	Professor
Seoul National University	Introduction to Bioinformatics	Bash introduction	Spring 2022	Asst. Prof. M. Steinegg (martin.steinegger@sn

How-to Guides

Setup

We recommend to use google cloud shell editor if you don't have any bash environment available. If you want to use your local environment, you can always use yours but **docker** must be pre-installed and available for setting dependencies.

About cloud shell editor, see CloudShellEditor section.

If you open the terminal, please clone this repository to your working directory and change the current directory to the cloned directory.

```
git clone <GITHUB_REPOSITORY_LINK>
cd ./<GITHUB_REPOSITORY_NAME>
```

After cloning, execute "setup.sh"

```
./setup.sh
```

If "setup.sh" executed well, you can see the username changed to "exercise01". The copied repository is at "~/exercise01" and please work within this directory.

Execution

Please write your commands to commandXX.sh and submit the whole repository by pushing the repository. You can run your all "commandXX.sh" files with "main.sh" or run independently.

- Run with main.sh

```
bash ./main.sh
```

- Run commandXX.sh independently

  ```
  bash ./command/commandXX.sh
  ```

Submission

To submit your result, follow these steps:

- Step 1. Write your commands in commandXX.sh in the "command" directory
- Step 2. Write the result to ./result/resultXX_X.txt or ./result/resultXX_X.csv

  > "." is relative path for current working directory and "./result/" is the directory named "result" in the current directory. For the first exercise, please copy the numbers or results from the terminal and paste it to result files. From the second exercise, you may use redirection to files.

- Step 3. Add edited files to git and commit. You can use the commands below or GUI of cloud shell editor.

  ```
  # Tell git who you are.
  git config --global user.email "you@example.com"
  git config --global user.name "Your Name"
  # Update the change.
  git add .
  git commit -m "Initial commit" # The message can be changed as you
  want.
  ```

- Step 4. Submit your answers by pushing the cloned repository.

  ```
  git push origin master
  ```

  You may need to enter your github ID and password in this step. If you have problems with using git, you can use the interface of github. You can edit commandXX.sh or resultXX.txt directly in the github repository web site as follows:

- Step 1. Click a file to edit and click the pencil icon at the top right corner. When you hover the cursor over the icon, you will see the phrase "Edit this file".
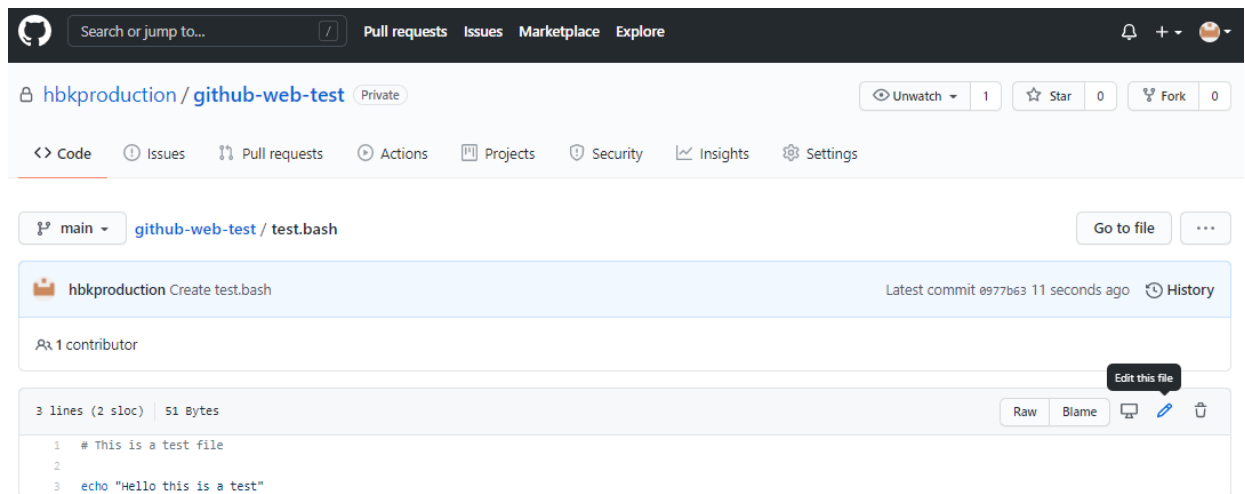
- Step 2. Edit the file with the text editor and click the green button with "Commit changes".
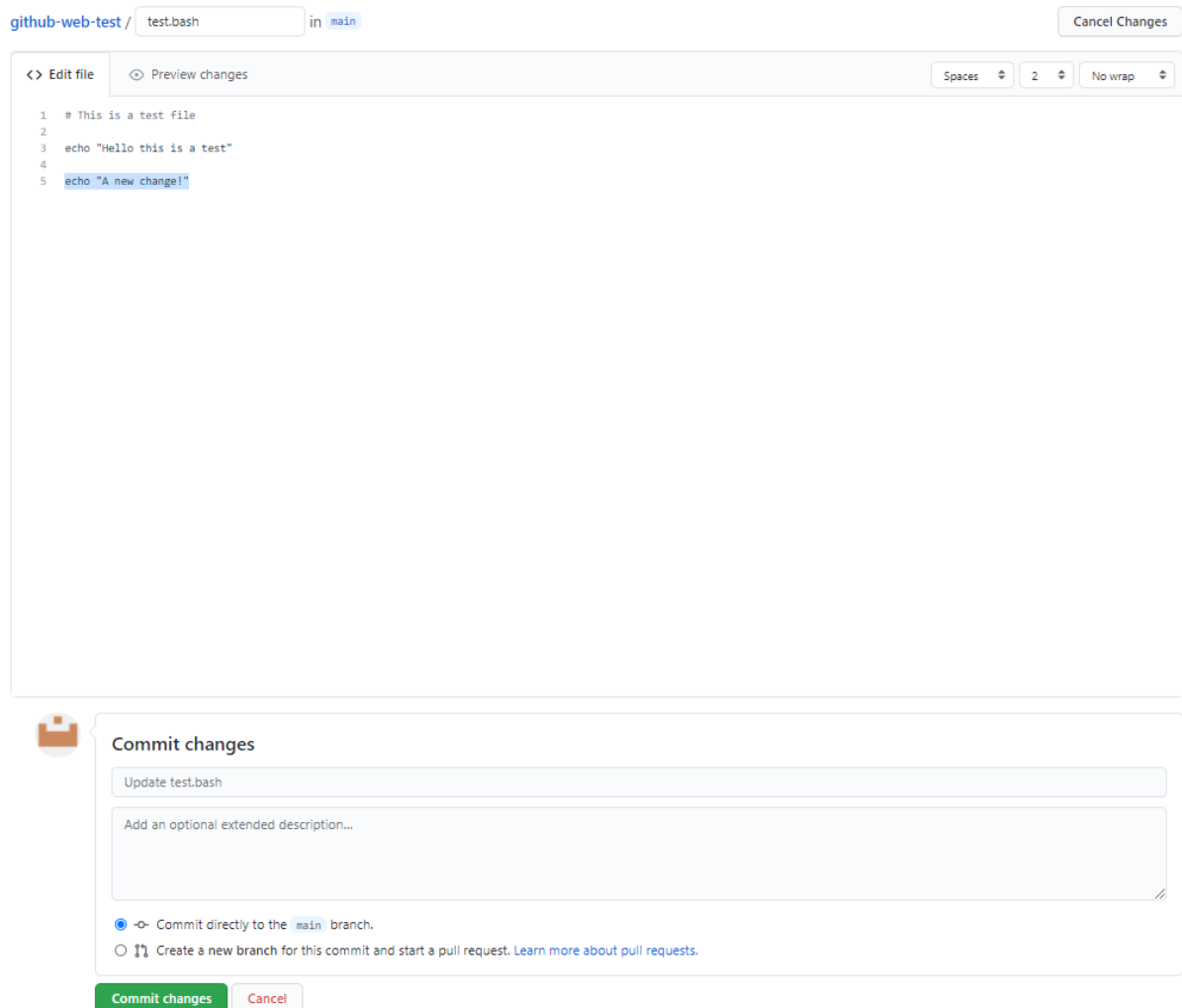
You can submit your result multiple times and the last version before the deadline will be graded.

# Exercise 01: Bash introduction #1

The goal of this exercise is to improve skills working with basic bash commands in UNIX environment.

> Commands to learn wget, gzip, mkdir, mv, cd, ls, grep, wc, head, cat, pipe "|" Please write your commands in commandXX.sh and save your result to resultXX.txt (or resultXX.csv)

command01.sh

1. Print your student number and github ID and copy the printed result to **student_id.txt**. The format should be like "2021-#####,harryporter".
   You can check your github ID at the top right corner of the github page.
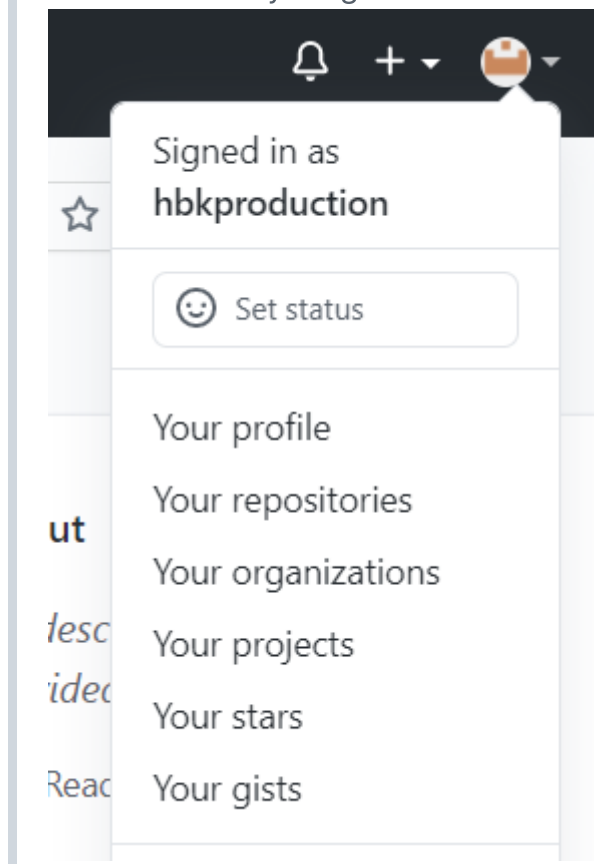
2. Download the fasta file with NCBI accession. The downloaded file should be named like "NC_XXXXXX.X.fasta" (No result file)

- Sequence name: Human herpesvirus 1 strain 17, complete genome
- NCBI accession number: NC_001806.2
- Download link:

- http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?
    db=nucleotide&id=NC_001806.2&rettype=fasta
- Hint
    - You can avoid parsing error by embracing the link with " ".
    - The name of the downloaded will be not proper : use -O option (search its usage)

command02.sh

1. Create a directory, named as the given sequence name, "Human_herpesvirus_1_strain_17", in data directory. Move the downloaded fasta file to the directory created. (No result file)

2. Print the list of files inside the data directory and copy the printed result to **result02.txt**.

command03.sh

1. Print first 5 lines of the downloaded fasta file and copy the printed result to **result03_1.txt**.
2. Print last 5 lines of the downloaded fasta file and copy the printed result to **result03_2.txt**.
3. Print from line 170 to line 174 (5 lines) of the downloaded fasta file and copy the printed result to **result03_3.txt**.

command04.sh

1. How many lines are in this fasta file except for the header line? (The header lines start with "#".) Print the number of lines except header lines and save the number to **result04_1.txt**.

2. How many lines of this fasta file 'not' containing "ATATA"? Print the number of lines with "ATATA" and save the number to **result04_2.txt.**

3. From the fasta file, print the first 5 lines containing "GGTTC" and copy the printed result to **result04_3.txt**.

command05.sh

1. Restriction enzymes cut DNA by acting only on certain motifs (motif means specific sequences of bases). EcoRI is a well-known restriction enzyme from *E. coli* and is used in various molecular genetics techniques. Identify the motif of EcoRI by searching online, print the number of motifs in the fasta file and save the number to **result05_1.txt**.

   > EcoRI motif consists of 6 bases. Don't consider the sites split by the new line character - just count the sites fully contained in a line.

2. Another restriction enzyme, HincII from *Haemophilus influenzae* recognizes a specific nucleotide pattern like this:
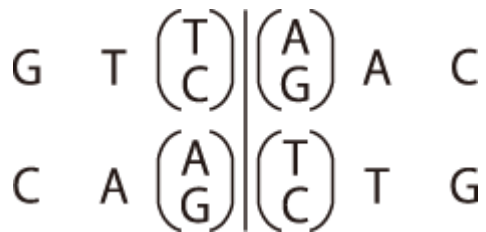
   Print the number of HincII recognition sites in the fasta file and save the number to **result05_2.txt**.

   > Don't consider the sites split by the new line character. Try using regular expressions.

3. Print the number of bases (A, T, G, C) in the fasta file and fill in the numbers of **result05_3.csv**.

CloudShellEditor

Guide document by Google is great: [link to guide](#).

You must have a Google account to use cloud shell (SNU account is fine). You can start to work on this exercise in Cloud shell as follows:

1. Open the link https://ide.cloud.google.com

2. Open a terminal from the cloud shell editor. You can open a new terminal by clicking the icon contains ">_" or selecting terminal at the menu.
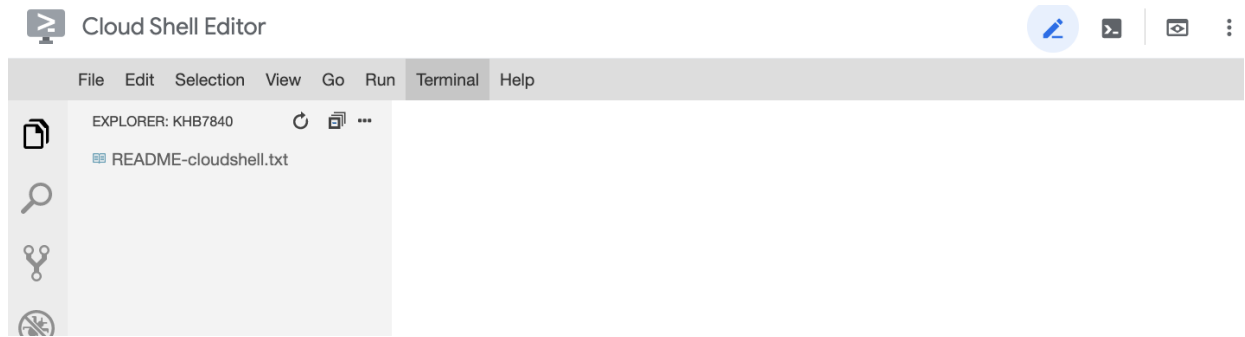
As a free service, Google Cloud Shell has some limitations. Please check these before doing your jobs.

Limitations

- It takes some time when initiating.
- 5GB storage limit
- Cannot be opened with secret mode of browser
- **If the session is inactive for 20 minutes, the instance is terminated and any modifications outside $HOME will be deleted.**
- Weekly usage limitation: ~50hrs
- If you don't access cloud shell for 120 days, the $HOME will be erased.

WARNING: Please don't provide your billing information unless you use google cloud platform.

If you have problems, please ask me at the zoom meeting or mail to me (jbeom@snu.ac.kr)