

Homework 3

Skylar Chan

2023-2-10 (due 2023-2-17)

Contents

| | | |
|----------|---|----------|
| 1 | Redirections | 1 |
| 2 | Subshells | 1 |
| 3 | Reverse complement of genetic sequence | 2 |
| 4 | File descriptors | 2 |
| 5 | Sort, Shuf, Uniq | 3 |

1 Redirections

I've put `mystery.sh` in the `public/hw3` directory where the course space is located. Make a copy of it in your home folder, and make it executable. Then you can run it. For example if you are in the same directory as your executable copy of `mystery.sh` you can just run `./mystery.sh` to run the script (the leading period is your current working directory, aka "here"). What will happen is that each line will be run as if you had typed it in.

1. Look at the line starting with `>`. What is the effect of this line, and why does that happen? Hint: what is the `stdin`?
2. Same thing as question 1 with the line starting with `»`.

2 Subshells

3. `diff` compares two files or two directories and reports the differences. For example, the following command reports if the directories are identical.

```
diff -r /:/ /cell_root/
```

Devise another way to take the `diff` of two directories using subshells. Hint: try an `ls` flag.

3 Reverse complement of genetic sequence

I put a sample of the *Arabidopsis thaliana* genome in the `public/hw3` folder. The file ends with `.fa`, and it's in the FASTA format, where each sequence starts with a `>` and the contents of the sequence follow.

4. `wc` counts words and lines. Run `wc` on the genome. What are the 3 numbers reported and what do they represent?
5. Devise a pipeline (a command with pipes) to get 100 lines from the middle of `athal` and save it to a file named `athal-sample.fa`. The file should only contain A,C,T,G.
6. `tr` can translate and delete characters. Give a command that uses `tr` to delete newlines from `athal-sample.fa` and saves the output to a file named `athal-sample-oneline.fa`. The newline character is `\n`.
7. `rev` reverses text. Devise a pipeline that takes the reverse complement of `athal-sample.fa`. You can use an online reverse complementer to check your command.
8. Give a command that combines the commands in questions 5, 6, and 7 using pipes. In other words, no saving the output to a file. The command should be one line.

4 File descriptors

9. `/dev/stdin`, `/dev/stdout`, and `/dev/stderr` are symlinks to file descriptors. What number file descriptor are they?
10. What happens when you run `echo y | tee /dev/stdin`? Why does the output look like that? Use `Ctrl+C` to exit.
11. Recall that `/dev/tty` prints to the terminal directly. What does the following command do? Explain each step of the pipeline.

```
echo y | tee /dev/tty | tee /dev/stdout | tee /dev/stderr | wc -l
```

12. **Extra Credit:** Suppose I chain a finite number of `tee /dev/stdout/` repeatedly like so:

```
echo y | tee /dev/stdout | tee /dev/stdout | tee /dev/stdout | ...
```

What formula expresses the number of lines that would be printed at the last `tee`? Explain your answer.

5 Sort, Shuf, Uniq

`shuf` shuffles files, `sort` sorts files, and `uniq` filters adjacent lines in the input.

13. What is the difference between `sort -u` and `sort | uniq`? When might you have to use one over the other?
14. **Extra Credit:** What is the difference between `sort -R` and `shuf`? When might you have to use one over the other?
15. Suppose you want to sort a file in-place (that is, replace the file with the sorted version). At first you might try `sort file > file`, but that won't work. What happens and why? What's the correct command to sort in-place?
16. `sort` has many options to sort differently depending on which sort is given. What do the following options do?
 - `-d`
 - `-g`
 - `-n`
 - `-h`
17. **Extra Credit:** Provide a file (at least 5 lines long) that will give a different result when sorted using these different switches. In other words, write a file that gives different results for:
 - `sort`
 - `sort -n`
 - `sort -V`