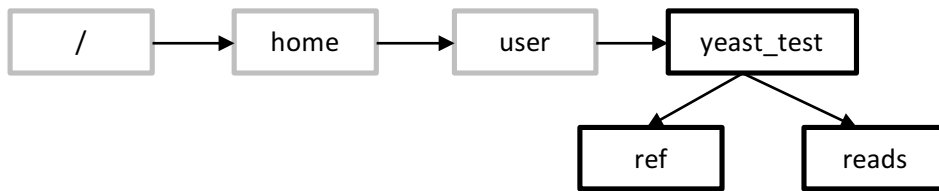


# Introduction to Linux: Exercises

- 1) Open a Terminal window.
- 2) Get the path of your working directory.
- 3) List the content of your home directory.
- 4) Create the following folder structure (The grey folders already exist).



- 5) Use `ls` to check if it worked. If you accidentally created a folder with an incorrect name use `rmdir` to remove it.
- 6) What are the commands `tar` and `curl/wget` for? Check the man pages or the internet.

- 7) Download the yeast read sequences into the “yeast\_test” folder.  
(URL: [http://www.cibiv.at/~philipp\\_/files/yeast-reads.tar](http://www.cibiv.at/~philipp_/files/yeast-reads.tar) )
- 8) Move the “yeast-reads.tar” files to the “reads” folder.
- 9) Open the man pages for the command “tar” to see how to unzip the archive containing the reads “yeast-reads.tar”.
- 10) Unzip “yeast-reads.tar”
- 11) List the contents of the “reads” folder.
- 12) If everything worked, delete the “yeast-reads.tar” file.
- 13) Combine the three fastq files into a single file “yeast-reads\_all.fq”.
- 14) Delete all fastq files except “yeast-reads\_all.fq”
- 15) Extract the first 10 sequences from file “yeast-reads\_all.fq” and save them as a new file called “test\_reads.fq”.  
*The file is a fastq file. Each sequence consists of four lines in the FASTQ format.*
- 16) Extract the last 5 sequences from the file “yeast-reads\_all.fq” and append them to the file “test\_reads.fq”.  
*To append data to a file use >> instead of >. The file is a fastq file. Each read consists of four lines in the FASTQ format.*
- 17) Open the file to see whether it worked. It should contain 15 sequences = 60 lines  
(command: cat, head, less, nano).
- 18) Use “nano” to remove the last sequence (that consist only of “N” characters).  
*Use Ctrl + K to remove a line. Use Ctrl + O to save and Ctrl + X to exit nano.*

- 19) Count how many sequences are contained in the files "yeast-reads\_all.fq" and "test\_reads.fq" (command: wc, Try to find the parameter that tells wc to only count the lines)
- The file is a fastq file. Each read consists of four lines in the FASTQ format.*
- 20) Download the yeast reference sequence and unzip it into the "ref" folder
- URL: [http://www.cibiv.at/~philipp\\_/files/yeast-ref.tar.gz](http://www.cibiv.at/~philipp_/files/yeast-ref.tar.gz)
- Use "tar -xvzf" for unzipping*
- 21) Count the number of sequences present in "yeast-ref.fq" (command: grep)
- The file is in fasta format. Each sequence entry starts with an ">"*
- 22) Extract the names of all reference sequences and save them in a new file "ref\_names.txt".
- 23) Open "ref\_names.txt" to see whether it worked.
- 24) Get the line number at which the sequence with the name "chromosome=IX" starts in the file "yeast-ref.fasta".
- Check the grep man page.*
- 25) Download the hello\_world.sh program in the "yeast\_test" folder and unzip it.
- URL: [http://www.cibiv.at/~philipp\\_/files/hello\\_world.tar.gz](http://www.cibiv.at/~philipp_/files/hello_world.tar.gz)
- Use the command "tar -xvzf" to unzip the hello\_world.tar.gz file*
- 26) Try to execute hello\_world.sh.
- 27) Make the file executable for the user and the group (command: chmod).
- 28) Run hello\_world.sh

#### Extra Exercises:

1. Use awk to extract only the read nucleotide sequences (the second line of every 4-line fastq entry) "yeast-reads\_all.fq" and save them as a new file "read\_sequences.txt".

*Hint: The command "awk '(NR+3) % 4 == 0' yeast-reads\_all.fq" extracts only the read names from the fastq file. It can be easily adapted to*

1. Use cut to extract the first 15 bases from each read sequence in "read\_sequences.txt" and combine the command with grep and wc -l (using pipes) to count the number of read sequences that contain "CAT" in their first 15 bases.

Hint: *The command "cut -c1-15" will extract the first 15 characters of each line of a file.*

1. Use cut and grep and wc -l to find out how many read sequences in "read\_sequences.txt" have an "A" as their 12th base.

Hint: *The command "cut -c12" will extract the 12th character of each line of a file.*