

Lesson 9: Read and write files in R

Modesto

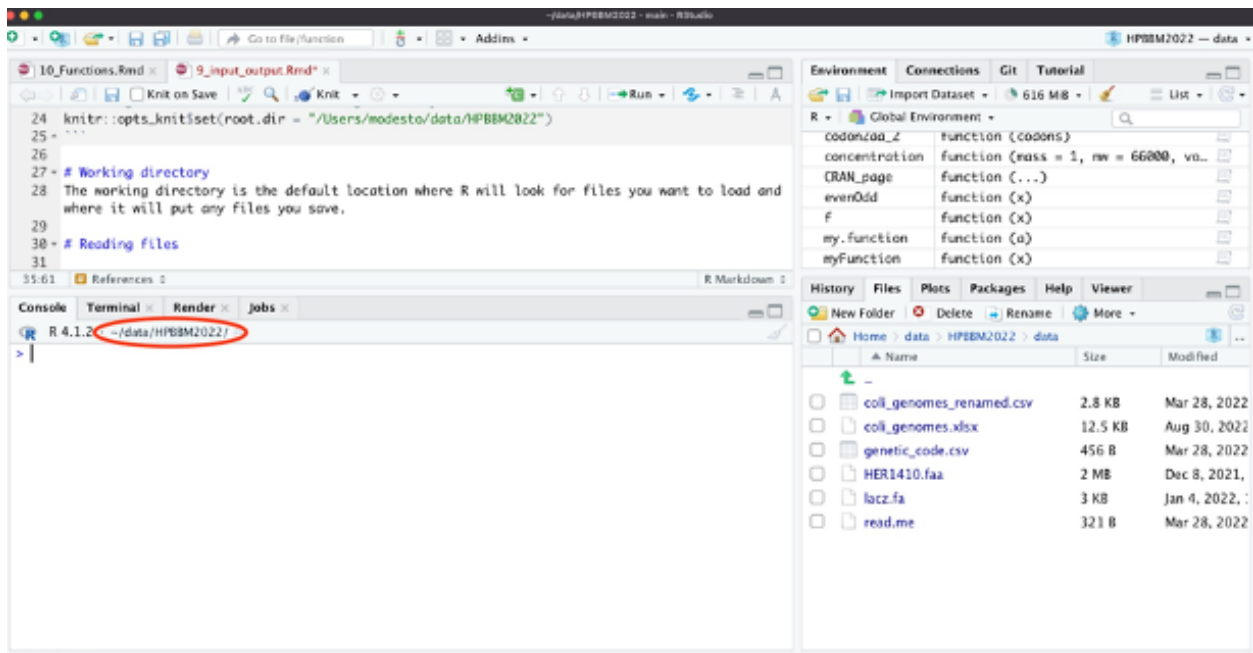
2022-08-30

Contents

Working directory	1
Reading files	2
References	2
Short exercises	2
Script-ing	3

Working directory

The working directory (wd) is the default location where R will look for files you want to load and where it will put any files you save.



Note that the abbreviation `~` stands for your home directory (for me `/Users/modesto` or `/home/modesto` on MacOS or Linux, respectively).

The functions `getwd()` and `setwd()` allow you to check and change the wd.

```
getwd()

## [1] "/Users/modesto/data/HPBBM2022"

# input of setwd() is a string -> requires quotes
setwd("/Users/modesto")
`?`(setwd())
getwd()

## [1] "/Users/modesto"

# you must write an existing path
setwd("/Users/mredrejo/data/HPBBM2022")

## Error in setwd("/Users/mredrejo/data/HPBBM2022"): no es posible cambiar el directorio de trabajo
setwd("/Users/modesto/data/HPBBM2022")
getwd()

## [1] "/Users/modesto/data/HPBBM2022"
```

In RStudio, the *default working directory* can be set from the “tools” and “global options” menu. Also, you can change the wd for your session in the menu Session>Set Working Directory and change it to that of source file (for instant your R script), the project or the selected directory in the files panel.

Additionally, RStudio encourages the use of Projects (see link below in References). RStudio Projects make it straightforward to divide your work into multiple contexts, each with their own working directory, workspace, history, and source documents. When you open a Project in RStudio, the current working directory is set to the project directory.

Reading files

References

- *An introduction to R*, <https://intro2r.com/work-d.html>
- *R programming for data science*, <https://bookdown.org/rdpeng/rprogdatascience/>
- Using RStudio projects, <https://support.rstudio.com/hc/en-us/articles/200526207>
-

Short exercises

1. Install the package `readxl` and read the file *coli_genomes.xls* in the *data* folder.
2. Export the dataframe from exercise 1 as a *csv* file. Read it as a new dataframe object. Is there any difference between the two dataframes?
3. Write a function called `micro()` that transforms concentrations units: molar (M) into micromolar (μ M)
4. Write a function that transform mass into concentration (in μ M and with four decimal digits). For simplicity, we can consider that the units of mass, molecular weight and volume are μ g, kg/mol and μ L, respectively.

5. Write a function that calculate your approximate age in months. Check the functions `date()`, `Sys.Date()`, `as.Date()`, and `difftime()`. See some examples [here](#) or [here](#).

Script-ing

Write a function that translate any nucleotide sequence into a protein sequence. Hints:

1. You may need to check for the number of nucleotides (i) and the number of codons ($j=i/3$) in the sequence.
2. You may use the function `substr()` to divide the sequence into codons. To do so, you may use a loop that split the nucleotides in groups of three (For instance, being the last nucleotide $j \times 3$ and the first $(j \times 3) - 2$).
3. You would add a `warning()` call when 1-2 nucleotides at the end of the sequence are not used.
4. Finally, use the package ***seqinr*** to read the fasta file *lacZ.fa*. Note that sequence objects are nested lists, thus, you'll need to extract the sequence as a plain text.