

Preparation for dada2 workflow workshop

To make sure we can run through the scripts with plenty of time to discuss the code, possibilities for downstream analysis and any changes that would be useful for you, please could you attempt to install the the key programs and packages used.

1. Install R, the latest version (v3.4.2) would be preferable, which can be downloaded [here](#)
2. Install RStudio. I will be using RStudio to demonstrate things. RStudio can be downloaded [here](#)
3. Install the packages used in the pipeline. I have attached a script that will hopefully install all the packages needed for the pipeline. Some of the packages (`phyloseq` and `dada2`) are hosted on [Bioconductor](#), which provides tools for analysis of genomic data in R. You will want to install the most up to date versions of each package, for Bioconductor this is using version 3.6. For people who have never used Bioconductor or R before, I imagine this will happen automatically. For others, at the end of the script there is code to update Bioconductor. Press `y` whenever prompted by the R console
4. Take a look at the GitHub repository [AB_dada2_pipeline_R](#). Download the folder and place somewhere on your laptop. The folder layout is needed
5. If you want to run the code on your own data, place your data in `data/raw_fastq` before the start of the workshop
6. Bring biscuits or cake. This is the most important prerequisite

Disclaimer: I have only ran this a few times so I am likely know not much more than you! But lets see how we get on...

NB Please email me on d.padfield@exeter.ac.uk if you have any problems, especially with Windows machines