HW6

Vicki Hertzberg

3/15/2017

Homework 6

On 3/29 we are going to use `dada2` to process demultiplexed, trimmed fastq files into something that can be use in `phyloseq` as an OTU table and analyzed to examine the microbiome. In order to follow along with this in class we ask that you prepare for it. Your homework is to show, in an rmd file on your repo that you subsequently output to pdf to turn into the Canvas site, that you have done this advance work. This takes some time, most of it just to have the machine process, hence the need to get it done before class.

Please document that you have installed the necessary packages and downloaded the appropriate files.

1. Document that you have installed `dada2` and `phyloseq`, as well as `shortread` and `ggplot2`. Use the installation guide at the dada2 site: (http://benjjneb.github.io/dada2/index.html)[http://benjjneb.github.io/dada2/index.html].

- Sufficient documentation would consist of the following R commands (in an R chunk, of course), one line per required package:

library(*package_name*); packageVersion("*package_name*");

- Except for `ggplot2`, these packages are not available through CRAN; rather you will have to go through bioconductor. See the `dada2` installation instructions for guidance. Note: you should already have `ggplot2`, but please document it as well.

2. Please document that you have downloaded THREE additional files as follows:

- the fastq files from the Mothur MiseqSOP walkthrough. Download the file from the `dada2` walkthrough page, the link marked "example data used in the Mother MiSeq SOP". Unzip this dataset and store it on your computer.

- the RDP training set from (https://zenodo.org/record/158955#.WMlZjxjGwo8)[https://zenodo.org/record/158955#.WMlZjxjGwo8]. Do NOT unzip it, but DO store it in the same directory as your miseq files.

- the RDP species assignmet set from the same site as the RDP training set. Again, do not unzip it but DO store it in the same directory as the miseq files.

- You can provide evidence that you have downloaded these files with the following commands (in an R chunk):

path <- "~/MiSeq_SOP" # CHANGE THIS to the directory containing the fastq files after unzipping fns <- list.files(path) fns

Due on 3/22 at 5 pm.