16s rRNA analysis workshop

Hena R. Ramay

2019-10-06

Contents

Chapter 1

Introduction

Our intention is to develop workshop material as we go along. For each day of the workshop, the basic material will be uploaded and more details will be added based on your questions and problems we encounter. So pleae ask as many questions are you can to help us in making this workshop better!!

1.1 Workshop Schedule

We will try to cover the following material in the course:

- Day1
  - 9:30 - 10:30 am Introductions and a presentation of the basic concepts
  - 10:30 - 10:45 am Break
  - 10:45 - 12 pm Installations of the required packages and data download
  - 12 - 1 pm Lunch break
  - 1 - 3 pm Reading data in and inspecting read quality
- Day2
  - 9:30 -10:30 am Filter and trim + learn error rates + Sample inference + Merge samples
  - 10:30 - 10:45 am Break
  - 10:45 - 12 Generate sequence table and remove Chimeras
  - 12 - 1 pm Lunch break
  - 1- 3 pm Taxomnomy explanation + assign your sequences. If time permits, make a phyloseq object
- Day3
  - 9-10:30 am Using a phyloseq object to calculate alpha and beta diversity

  - 10:30 onwards we will use a real world dataset to re-do what we have learned here

1.2 Important links

- DADA2 Tutorial : link
- Day One presentaton : link
- Day One dataset : link

Chapter 2

Day One

We are very excited to teach this course for the first time and share what we know with you all

So lets start by talking about the very basics:

2.1 What are we trying to achieve

Our goal is very similar to gathering data on a city neighbourhood to find out who lives there, how the demographic changes over time or in case of a drastic event. We can gather more information by asking about neighbours, quality of life etc. Similarly when we are looking at microbial communities our first question is who is there, how abundant and how their presence changes over time or when conditions change. We can also ask questions like how the microbiomes are interacting with each other (metabolites).

For the scope of this workshop we will stick to the simple questions: who and how much?

2.2 Basics & Background

Here is the link to the lecture we will start with today: workshop

Key points are:

- Think of a hypothesis before doing an experiment
- Spend time on experiment design.
  - Sample size, 16s region to amplify etc
  - Talk to a bioinformatician

  - Think about the depth of sequencing if you want to capture the less abundant taxa
  - Add negative control to account for contamination
- Thoughtful data analysis is critical for successful identification of microbes

"If you torture the data long enough, it will confess."- Ronald Coase, Economist

2.3 DADA2 pipeline (v1.2)

From now on, we will be working on the DADA2 package version 1.12. DADA2 has great documentation and an excellent tutorial online that we will use to understand the pipeline. Please go to the following link http://benjjneb.github.io/dada2/tutorial.html

2.3.1 Data for the tutorial

The data to use for the tutorial can be downloaded from here

2.3.2 Getting ready (load packages and get file list)

functions that we will be using here are :

- list.files()
- sort()
- strsplit()
- basename()
- sapply()

2.4 Inspect read quality profiles

Chapter 3

Day Two

We describe our methods in this chapter.

Chapter 4

Day Three

Chapter 5

Summary