

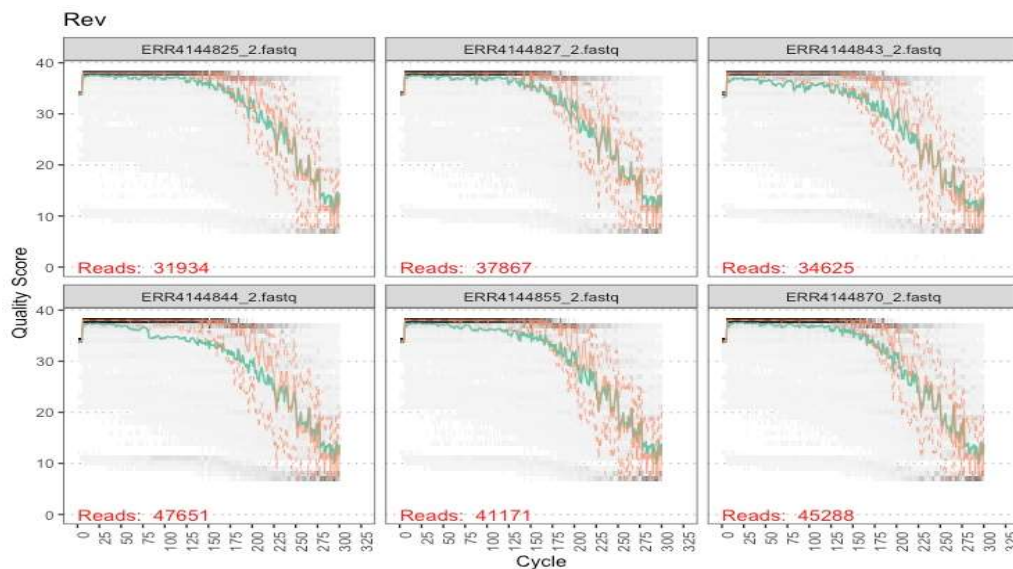
Ciclo di Training courses SUS-MIRRI.IT

Titolo “Analisi di dati metatassonomici: un flusso di lavoro in R”

Organizzatore: Università degli Studi della Basilicata

Periodo di erogazione: 13 e 15 Giugno 2023

Corso online



Iscrizione: <https://forms.gle/u7FSXKEXWT7zUteKA>

Docenti: Prof. Eugenio Parente

PROGRAMMA

Giorno 1: 13/06/2023 – ore 15:00-17:00

Analisi di dati metatassonomici: un flusso di lavoro in R – 1

In questo primo seminario verrà presentato brevemente (dopo una micro introduzione su R e RStudio) un flusso di lavoro per l'analisi di dati metatassonomici a partire da sequenze in formato fastq. Nella prima lezione ci si concentrerà su importazione di dati, applicazione di filtri di qualità e inferenza delle Amplicon Sequence Variants con DADA2.

Giorno 2: 15/06/2023 – ore 16:00-18:00

Analisi di dati metatassonomici: un flusso di lavoro in R – 2

In questo secondo seminario verrà presentata la seconda parte di un flusso di lavoro per analisi metatassonomiche con R, basato su DADA2.

Si partirà dalla costruzione della tabella di sequenze, per poi arrivare all'insieme delle tabelle dei campioni, dell'abbondanza delle ASV e dei taxa, oltre che dell'albero filogenetico, combinati in un oggetto phyloseq. Se sarà possibile verranno indicati brevemente approcci e pacchetti per l'ulteriore analisi statistica e grafica dei dati.



Informazioni e Iscrizioni: Per le informazioni sul corso inviare una e-mail al seguente indirizzo: teresa.zotta@unibas.it

Numero massimo di partecipanti: 100 partecipanti

Requisiti

Per partecipare con profitto al corso è necessaria almeno una conoscenza elementare di R e della IDE di RStudio e disporre di un notebook o personal computer (Unix, MacOS o Windows) con almeno 8 GB di RAM e con processori adeguati (possibilmente quad core, intel i5 o superiore o chip equivalente)

R (è necessaria la versione 3.3 o superiore, è preferibile scaricare la versione più recente) può essere scaricato qui: <https://cran.r-project.org/>¹

RStudio può essere scaricato qui: <https://posit.co/download/rstudio-desktop/>

E' necessario inoltre scaricare i database tassonomici di riferimento da qui: <https://benjjneb.github.io/dada2/training.html>²

I flussi di lavoro possono essere scaricati da qui³:

https://github.com/ep142/FoodMicrobionet/tree/master/dada2_pipeline

https://github.com/ep142/FoodMicrobionet/tree/master/dada2_pipeline_big_data

Il flusso di lavoro è basato su DADA2, con modificazioni:

<https://benjjneb.github.io/dada2/index.html>

E' caldamente consigliata la lettura almeno dei seguenti articoli:

- Callahan, B.J., Sankaran, K., Fukuyama, J.A., McMurdie, P.J., Holmes, S.P., 2016. Bioconductor Workflow for Microbiome Data Analysis: from raw reads to community analyses. *F1000Research* 5, 1492. <https://doi.org/10.12688/f1000research.8986.2>
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods* 13, 581–583. <https://doi.org/10.1038/nmeth.3869>
- Wen, T., Niu, G., Chen, T., Shen, Q., Yuan, J., Liu, Y.-X., 2023. The best practice for microbiome analysis using R. *Protein Cell*. <https://doi.org/10.1093/procel/pwad024>
- Bokulich, N.A., Ziemski, M., Robeson, M.S., Kaeher, B.D., 2020. Measuring the microbiome: Best practices for developing and benchmarking microbiomics methods. *Comput Struct Biotechnology J* 18, 4048–4062. <https://doi.org/10.1016/j.csbj.2020.11.049>
- Pollock, J., Glendinning, L., Wisedchanwet, T., Watson, M., 2018. The Madness of Microbiome: Attempting To Find Consensus “Best Practice” for 16S Microbiome Studies. *Applied and Environmental Microbiology* 84, 3225. <https://doi.org/10.1128/aem.02627-17>

¹ Se si possiede un Mac è meglio scaricare la versione per processori Intel; dovrebbe funzionare anche quella per arm64 ma non la ho ancora provata

² Lavoreremo esclusivamente con SILVA

³ Per scaricare inserire i path qui: <https://download-directory.github.io>

E' preferibile eseguire l'installazione di almeno una parte pacchetti necessari prima del corso usando questi comandi⁴:

```
# Install/load packages -----  
  
.cran_packages <- c("tidyverse", "parallel", "beep", "tictoc")  
.bioc_packages <- c("BiocManager","dada2", "phyloseq", "DECIPHER", "phangorn",  
  "BiocStyle", "ShortRead")  
  
.inst <- .bioc_packages %in% installed.packages()  
if(any(!.inst)) {  
  if(!.inst) {  
    install.packages("BiocManager")  
    .inst <- .bioc_packages %in% installed.packages()  
  }  
  if(any(!.inst[length(.inst)])) {  
    BiocManager::install(.bioc_packages[!.inst], ask = F)  
  }  
}  
  
.inst <- .cran_packages %in% installed.packages()  
if(any(!.inst)) {  
  install.packages(.cran_packages[!.inst])  
}  
# Load packages into session, and print package version  
sapply(c(.cran_packages, .bioc_packages), require, character.only = TRUE)  
# end script -----
```

Ulteriori pacchetti verranno installati in seguito.

⁴ Da eseguire in uno script nel source pane di RStudio