DATA PROCESSING FOR ILLUMINA 16S READS

Created/updated by: Laura Rossi **Date: May21/21**
Surette Lab, McMaster University
Hamilton, ON, Canada
www.surettelab.ca

BACKGROUND

- This outline is intended to be a guide to process raw reads, ie. fastq files, through the read processing pipeline DADA2. This outline is specific for data obtained through the Surette Lab and the Farncombe Metagenomics Facility. If your sequencing data is from somewhere else some specifics may differ.

PROGRAMS

- Cutadapt: https://cutadapt.readthedocs.io/en/stable/ or
- Trimmomatic: http://www.usadellab.org/cms/?page=trimmomatic
- DADA2: https://benjjneb.github.io/dada2/tutorial.html

PROTOCOL

- Primers **must** be trimmed for DADA2 to work correctly
    - V3 data from a 2x250bp run, trim both 5' and 3' ends of R1 and R2 using paired end functionality
        - R1-5'- ATTACCGCGGCTGCTGG
        - R2-5'- CCTACGGGNGGCWGCAG
        - R1-3'- CTGCWGCCNCCCGTAGG
        - R2-3'- CCAGCAGCCGCGGTAAT
    - V34 data from a 2x300bp run, trim only 5' end of R1 and R2
        - R1-5'- GGACTACNVGGGTWTCTAAT
        - R2-5'- CCTACGGGNGGCWGCAG
- Run DADA2
    - filterAndTrim parameters needed for the sequencing you have:
        - v3 data from a 2x250bp run
            - truncLen=c(130,130), maxN=0, maxEE=c(2,2)
        - v34 data from a 2x300bp run

        - truncLen=c(240,210), maxN=0, maxEE=c(2,2)
    - assignTaxonomy function used with tryRC=TRUE and the SILVA database (previously downloaded from the DADA2 website)

PRIMERS

v3f_341f – CCTACGGGNGGCWGCAG – as per Illumina protocol
v4r_806r – GGACTACNVGGGTWTCTAAT – as per Earth Microbiome protocol

REFERENCES

- MARTIN, Marcel. Cutadapt removes adapter sequences from high-throughput sequencing reads. **EMBnet.journal**, [S.l.], v. 17, n. 1, p. pp. 10-12, may 2011. ISSN 2226-6089. Available at: <http://journal.embnet.org/index.php/embnetjournal/article/view/200>. Date accessed: 01 june 2021. doi:https://doi.org/10.14806/ej.17.1.200.

- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina Sequence Data. *Bioinformatics*, btu170.

- Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." ***Nature Methods***, **13**, 581-583. doi: 10.1038/nmeth.3869.

APPENDIX A – FOR ADAPTERS AND BARCODES

PRIMERS: 16S RRNA V3 REGION REVERSE PRIMER

Primer	Adaptor	Sequence
V3_Rmod2	P5	aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctNNNNATTACCGCGGCTGCTGG

PRIMERS: 16S RRNA V3 REGION FORWARD BARCODED PRIMER

Primer	Adaptor	Sequence
V3_1F	P7	caagcagaagacggcatacgagat**CGTGAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_2F	P7	caagcagaagacggcatacgagat**ACATCG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_3F	P7	caagcagaagacggcatacgagat**GCCTAA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_4F	P7	caagcagaagacggcatacgagat**TGGTCA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_5F	P7	caagcagaagacggcatacgagat**CACTGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_6F	P7	caagcagaagacggcatacgagat**ATTGGC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_7F	P7	caagcagaagacggcatacgagat**GATCTG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_8F	P7	caagcagaagacggcatacgagat**TCAAGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_9F	P7	caagcagaagacggcatacgagat**CTGATC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_10F	P7	caagcagaagacggcatacgagat**AAGCTA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_11F	P7	caagcagaagacggcatacgagat**GTAGCC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG

V3_12F	P7	caagcagaagacggcatacgagat**TACAAG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_13F	P7	caagcagaagacggcatacgagat**CGTACT**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_14F	P7	caagcagaagacggcatacgagat**GACTGA**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_15F	P7	caagcagaagacggcatacgagat**GCTCAA**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_16F	P7	caagcagaagacggcatacgagat**TCGCTT**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_17F	P7	caagcagaagacggcatacgagat**TGAGGA**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_18F	P7	caagcagaagacggcatacgagat**ACAACC**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_19F	P7	caagcagaagacggcatacgagat**ACCTCA**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_20F	P7	caagcagaagacggcatacgagat**ACGGTA**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_21F	P7	caagcagaagacggcatacgagat**AGTTGG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_22F	P7	caagcagaagacggcatacgagat**CTCTCT**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_23F	P7	caagcagaagacggcatacgagat**CAAGTG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_24F	P7	caagcagaagacggcatacgagat**CCTTGA**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_25F	P7	caagcagaagacggcatacgagat**ACCACT**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_26F	P7	caagcagaagacggcatacgagat**AGTGTC**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_27F	P7	caagcagaagacggcatacgagat**AGAAGG**gtgactggagttcagacgtgtgctcttccgatctCCTACG

		GGAGGCAGCAG
V3_28F	P7	caagcagaagacggcatacgagat**TTATCC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_29F	P7	caagcagaagacggcatacgagat**TTAAGG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_30F	P7	caagcagaagacggcatacgagat**TTCTTG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_31F	P7	caagcagaagacggcatacgagat**TTCAAC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_32F	P7	caagcagaagacggcatacgagat**TTGTGA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_33F	P7	caagcagaagacggcatacgagat**TTGACT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_34F	P7	caagcagaagacggcatacgagat**TATTCG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_35F	P7	caagcagaagacggcatacgagat**TATAGC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_36F	P7	caagcagaagacggcatacgagat**TAACTC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_37F	P7	caagcagaagacggcatacgagat**TACCAA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_38F	P7	caagcagaagacggcatacgagat**TACGTT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_39F	P7	caagcagaagacggcatacgagat**TAGTAC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_40F	P7	caagcagaagacggcatacgagat**TAGATG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_41F	P7	caagcagaagacggcatacgagat**TCTACA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_42F	P7	caagcagaagacggcatacgagat**TCTGAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG

V3_43F	P7	caagcagaagacggcatacgagat**TCATGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_44F	P7	caagcagaagacggcatacgagat**TGTCTA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_45F	P7	caagcagaagacggcatacgagat**ATTCTC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_46F	P7	caagcagaagacggcatacgagat**ATTGAG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_47F	P7	caagcagaagacggcatacgagat**ATACCT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_48F	P7	caagcagaagacggcatacgagat**ATGCAA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_49F	P7	caagcagaagacggcatacgagat**AATCCA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_50F	P7	caagcagaagacggcatacgagat**AATGGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_51F	P7	caagcagaagacggcatacgagat**AACTAG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_52F	P7	caagcagaagacggcatacgagat**AACACT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_53F	P7	caagcagaagacggcatacgagat**AAGAGA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_54F	P7	caagcagaagacggcatacgagat**ACTTAC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_55F	P7	caagcagaagacggcatacgagat**ACATTG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_56F	P7	caagcagaagacggcatacgagat**ACGAAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_57F	P7	caagcagaagacggcatacgagat**AGTCAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_58F	P7	caagcagaagacggcatacgagat**AGAAGT**gtgactggagttcagacgtgtgctcttccgatctCCTACG

		GGAGGCAGCAG
V3_59F	P7	caagcagaagacggcatacgagat**CTTATG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_60F	P7	caagcagaagacggcatacgagat**CTAGAA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_61F	P7	caagcagaagacggcatacgagat**CATCTT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_62F	P7	caagcagaagacggcatacgagat**CACATA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_63F	P7	caagcagaagacggcatacgagat**CCAATT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_64F	P7	caagcagaagacggcatacgagat**CGATTA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_65F	P7	caagcagaagacggcatacgagat**GTTAGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_66F	P7	caagcagaagacggcatacgagat**GTAACA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_67F	P7	caagcagaagacggcatacgagat**GTGTAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_68F	P7	caagcagaagacggcatacgagat**GATAAG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_69F	P7	caagcagaagacggcatacgagat**GAATCT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_70F	P7	caagcagaagacggcatacgagat**TTCCGT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_71F	P7	caagcagaagacggcatacgagat**TTCGCA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_72F	P7	caagcagaagacggcatacgagat**TTGGTC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_73F	P7	caagcagaagacggcatacgagat**TGACAG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG

V3_74F	P7	caagcagaagacggcatacgagat**ATCTGC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_75F	P7	caagcagaagacggcatacgagat**ACACGA**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_76F	P7	caagcagaagacggcatacgagat**AGGTTC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_77F	P7	caagcagaagacggcatacgagat**CATGAC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_78F	P7	caagcagaagacggcatacgagat**GCTATC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_79F	P7	caagcagaagacggcatacgagat**GGACTT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_80F	P7	caagcagaagacggcatacgagat**GGCAAT**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_81F	P7	caagcagaagacggcatacgagat**TCTCGG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_82F	P7	caagcagaagacggcatacgagat**TCAGCG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_83F	P7	caagcagaagacggcatacgagat**TGTGCC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_84F	P7	caagcagaagacggcatacgagat**TGCACG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_85F	P7	caagcagaagacggcatacgagat**AAGGCC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_86F	P7	caagcagaagacggcatacgagat**ACCAGG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_87F	P7	caagcagaagacggcatacgagat**AGCCTG**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_88F	P7	caagcagaagacggcatacgagat**AGCGAC**gtgactggagttcagacgtgtgctcttccgatctCCTACGGGAGGCAGCAG
V3_89F	P7	caagcagaagacggcatacgagat**CTACGC**gtgactggagttcagacgtgtgctcttccgatctCCTACGG

		GAGGCAGCAG
V3_90F	P7	caagcagaagacggcatacgagat**CTCCAG**gtgactggagttcagacgtgtgctcttccgatctCCTACGG GAGGCAGCAG
V3_91F	P7	caagcagaagacggcatacgagat**CCGTAG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_92F	P7	caagcagaagacggcatacgagat**CGGTGT**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_93F	P7	caagcagaagacggcatacgagat**CGGAAC**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_94F	P7	caagcagaagacggcatacgagat**GTGCTG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_95F	P7	caagcagaagacggcatacgagat**GAACGG**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG
V3_96F	P7	caagcagaagacggcatacgagat**GGATGC**gtgactggagttcagacgtgtgctcttccgatctCCTACG GGAGGCAGCAG