Cite: Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods. 2016;13:581–3.

library(dada2)

path <- "C:/path/to/your/trimmed/reads/folder"

list.files(path)

# Forward and reverse fastq filenames have format: SAMPLENAME\_R1\_001.fastq and SAMPLENAME\_R2\_001.fastq

fnFs <- sort(list.files(path, pattern="\_R1\_val\_1.fq", full.names = TRUE)) #pattern can be modify

fnRs <- sort(list.files(path, pattern="\_R2\_val\_2.fq", full.names = TRUE)) #pattern can be modify

if(length(fnFs) != length(fnRs)) stop("Forward and reverse files do not mach")

# Extract sample names, assuming filenames have format: SAMPLENAME\_XXX.fastq

sample.names <- sapply(strsplit(basename(fnFs), "\_"), `[`, 1)

# Place filtered files in filtered/ subdirectory

filtFs <- file.path(path, "filtered", paste0(sample.names, "\_F\_filt.fastq.gz"))

filtRs <- file.path(path, "filtered", paste0(sample.names, "\_R\_filt.fastq.gz"))

#Set truncLen and minLen according to your dataset

#AOB amoA assembly: truncLen=c(229,229), minLen = 229

#AOA amoA gap: truncLen=c(200,200), minLen = 200

out <- filterAndTrim(fnFs, filtFs, fnRs, filtRs, truncLen=c(229,229), minLen = 229,

maxN=0, maxEE=c(2,2), truncQ=2, rm.phix=TRUE,

compress=TRUE, multithread=FALSE) # On Windows set multithread=FALSE

out