# bacteria spades assembly

by Yanyan Liu

## what for

This is the nextflow pipeline that does bacteria genome assembly using spades. The assembly metrics are collected by QUAST. BUSCO metrics is also produced. It is currently does not have the downsample option.

## how to run

### 1.copy the code below to nextflow\_spades.sh

#!/bin/bash  
  
run=20250627\_NextSeq2000  
plates=Ecoli\_\*\_FASTQ  
sample\_map=ecoli\_samples\_map.csv  
analysis=Ecoli\_spades\_busco  
  
  
  
/software/nextflow-align/nextflow run \  
/software/nextflow-spades-quast/main.nf \  
-work-dir s3://seqwell-analysis/${run}/${analysis}/work \  
--analysis $analysis \  
--run $run \  
--plates $plates \  
--sample\_map $sample\_map \  
-bg -resume

### 2.run as *bash nextflow\_spades.sh*

### 3.sample map example and requirement

sample\_id,ref  
Ecoli\_Std\_A04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_B04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_C04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_D04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_E04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_F04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_G04,Ecoli\_SAMN07731009\_gcf  
Ecoli\_Std\_H04,Ecoli\_SAMN07731009\_gcf

* The sample map needs two columns, with the right header: *sample\_id* and *ref*
* For sample\_id, it needs to be matched to the fastq file names. For example, if the fastq file is Ecoli\_HS\_H07\_R1\_001.fastq.gz and Ecoli\_HS\_H07\_R2\_001.fastq.gz. The sample\_id will be *Ecoli\_HS\_H07*