bacteria spades assembly

date: 20260108

by Yanyan Liu

what for

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

how to run

1.copy the code below to nextflow_spades.sh

```
#!/bin/bash

run=20251218_MiSeqi100-Morty
plates=REL_25J-B_FASTQ
sample_map=ecoli_map_run_20251218_MiSeqi100_Morty.csv
analysis=Ecoli_spades_test



/software/nextflow-align/nextflow run \
/software/nextflow-assembly-spades/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
REL_25J-B_A08,ecoli_REL606
REL_25J-B_B08,ecoli_REL606
REL_25J-B_C08,ecoli_REL606
REL_25J-B_D08,ecoli_REL606
REL_25J-B_E08,ecoli_REL606
REL_25J-B_F08,ecoli_REL606
REL_25J-B_G08,ecoli_REL606
REL_25J-B_H08,ecoli_REL606
```

- 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 REL_25J-B_A08_R1_001.fastq.gz and REL_25J-B_A08_R2_001.fastq.gz. The sample_id will be *REL_25J-B_A08*