Tutorial for running AMRplusplus on the MSI server

Please contact Noelle Noyes or Enrique Doster with any questions.

1. Brief overview
   1. In short, you have to set up your account with the right tools, locate sample files, identify which pipeline you want to run, pick location of output files, and finally run the pipeline using “screen”.
   2. This is an example command to run the pipeline:

nextflow run main\_amr\_plus\_plus\_v1.nf -resume -profile local\_MSI --threads 10

--reads '/home/noyes046/shared/projects/resistome\_mobilome\_PlosONE/raw\_reads/draxxin/\*\_R{1,2}\_001.fastq.gz'

--host /panfs/roc/risdb/genomes/Bos\_taurus/Bos\_taurus\_UMD\_3.1/bwa/Bos\_taurus\_UMD\_3.1.fa

--amr /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_database\_v1.02.fasta

--annotation /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_annotations\_v1.02.csv

--kraken\_db /home/noyes046/shared/databases/kraken2\_databases/Rumen\_kraken\_v2\_July2019

-w */scratch.global/TEMP\_DIR*

--output */scratch.global/*OUTPUT\_DIR

1. Setting up your account (only needs to be done once)
   1. Before running AmrPlusPlus, you need to make sure that your computing environment is set up. The flag “-profile” provides several options depending on what is most convenient.
      1. -profile singularity
         1. This is the most convenient option if you can install Singularity on your computing system. This way, a singularity container will be downloaded and provides easy access to all pipeline dependencies. <https://github.com/meglab-metagenomics/amrplusplus_v2/blob/master/docs/dependencies.md>
      2. -profile local
         1. This profile can be edited to best suit your computing environment.
      3. -profile local\_angus
         1. Example profile used for MEG’s computing cluster, “angus”
2. Locate sample files
   1. First identify which samples you will run and their location. This will make up the “--reads” flag for the pipeline. Use regular expressions to specify the forward and reverse reads.

--reads '/home/noyes046/shared/projects/resistome\_mobilome\_PlosONE/raw\_reads/draxxin/\*\_R{1,2}\_001.fastq.gz'

1. Identify which pipeline version you want to run
   1. Besides the original AMRplusplus pipeline which includes both microbiome and resistome characterization, we have different options available depending on your goal.
      1. For this example we’ll use the full AMRplusplus pipeline. For the full pipeline we’ll need to specify the following flags:
         1. -- amr
         2. -- amr\_annotations
         3. -- kraken\_db
         4. -- host
      2. Which version of megares do you want to run? If you are not using megares, some of the output will not work correctly (e.g. rarefaction tables). These are the standard commands:

--amr /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_database\_v1.02.fasta

--annotation /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_annotations\_v1.02.csv

* + 1. Do you need a kraken database for microbiome classification?
       1. We recommend using our modified kraken database which contains reference genomes for bacteria, archaea, protists, and an additional 913 un-culturable bacteria found in the bovine rumen:

--kraken\_db /home/noyes046/shared/databases/kraken2\_databases/Rumen\_kraken\_v2\_July2019

* + 1. Which host DNA do you need to remove?

--host /panfs/roc/risdb/genomes/Bos\_taurus/Bos\_taurus\_UMD\_3.1/bwa/Bos\_taurus\_UMD\_3.1.fa

* + 1. Do you need to change the Trimmomatic parameters for quality trimming of reads?
       1. AMRplusplus has standard parameters for quality trimming that you can see in the “nextflow.config” file within the AMRplusplus repository. You can change those parameters in the file directly or add them to your nextflow command like the following example:

--adapters /panfs/roc/msisoft/trimmomatic/0.33/adapters/all\_illumina\_adapters.fa

1. Pick a location for your output files (and temporary files)
   1. Choosing the location for your output files is VERY important and can cause issues if you choose a location without enough space.
   2. For large projects, please check the status of the Noyes’ server to ensure there will be enough storage space for the output. If in doubt, we recommend using MSI’s scratch space which has a storage capacity of >700TB, but you must remember to delete the temporary files after running the pipeline and moving your output to the Noyes server. Please note that files on the scratch space are automatically deleted after 30 days.
   3. For example, the temporary files can be output here:

-w */scratch.global/TEMP\_DIR*

* 1. And your output files can be directed like this:

--output */scratch.global/*OUTPUT\_DIR

1. Running AMRplusplus
   1. Log into MSI
      1. $ ssh edoster@login.msi.umn.edu
   2. Log into MESABI
      1. $ ssh mesabi
   3. Log into the Noyes server’s computation node
      1. $ ssh cn4201
   4. Create a “screen” which allows you to run the pipeline in the background.
      1. Create a new screen named “test\_run”
         1. $ screen -RD test\_run
      2. Detach from that screen
         1. click the Ctrl + A + D buttons on your keyboard
      3. See which screens you have created
         1. $ screen -ls
      4. Re-attach to the test\_run screen
         1. $ screen -x test\_run
   5. Navigate to the directory with the AMRplusplus pipeline
      1. $ cd /home/noyes046/shared/bioinformatic-nextflow-pipelines
   6. While still “attached” to your screen, you’ll enter the full command with all of the various components. The first part will remain the same and you just need to update the name of the pipeline you are running (file ending in .nf):

nextflow run main\_amr\_plus\_plus\_v1.nf -resume -profile local\_MSI --threads 10

--reads '/home/noyes046/shared/projects/resistome\_mobilome\_PlosONE/raw\_reads/draxxin/\*\_R{1,2}\_001.fastq.gz'

--host /panfs/roc/risdb/genomes/Bos\_taurus/Bos\_taurus\_UMD\_3.1/bwa/Bos\_taurus\_UMD\_3.1.fa

--amr /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_database\_v1.02.fasta

--annotation /home/noyes046/shared/bioinformatic-nextflow-pipelines/containers/data/amr/megares\_annotations\_v1.02.csv

--kraken\_db /home/noyes046/shared/databases/kraken2\_databases/Rumen\_kraken\_v2\_July2019

-w */scratch.global/TEMP\_DIR*

--output */scratch.global/*OUTPUT\_DIR