Salmon on HPC cluster:

What is Salmon?

Salmon is a wicked-fast program to produce highly accurate, transcript-level quantification estimates from RNA-seq data. Salmon achieves its accuracy and speed via several different innovations, including the use of *selective-alignment* (accurate but fast-to-compute proxies for traditional read alignments), and massively parallel stochastic collapsed variational inference. The result is a versatile tool that fits nicely into many different pipelines. For example, you can choose to make use of our *selective-alignment* algorithm by providing Salmon with raw sequencing reads, or, if it is more convenient, you can provide Salmon with regular alignments (e.g., an unsorted BAM file with alignments to the transcriptome produced with your favorite aligner), and it will use the same wicked-fast, state-of-the-art inference algorithm to estimate transcript-level abundances for your experiment.

The documentation for the Salmon is found in the following links:

Documentation

GITHUB

Versions Available:

Salmon– 0.11.3

Salmon– 1.3.0

How to load a version of Salmon?

To load a version of Salmon on the HPC, use the following command:

module avail bio/salmon

The version will be listed. To use a version of software, use following command:

module load bio/salmon/1.3.0

Verify by using this command:

module list

The loaded software and dependencies will be shown.

How to use Salmon on the cluster?

Here is one example of how to use Salmon on the cluster. To get a test file to process through Salmon, use the following commands

```
cd ~ && mkdir Salmon_test && cd Salmon_test       # Make a test directory on home #fo
lder


# Copy the test files from Salmon source folder

cp /share/apps/salmon/salmon-1.3.0_linux_x86_64/sample_data/* ~/Salmon_test/


# Now after you load your module and copied test files, we can make a script to run S
almon
```

Interactive method:

To run the program interactively, follow the steps:

```
#Open a bash session on compute node

srun -p main --qos main -n 1 -c 12 --mem 10G --pty bash


#load the module

module load bio/salmon/1.3.0
```

```
# start your commands here
# This method is ideal for short runs which produces runtime output
```

The Script:

```
#!/bin/bash
$SBATCH -J Jobname # Jobname
#SBATCH -n 1       # Nodes per task
#SBATCH –mem=10G    # Must need to allocate memory to run Salmon– at least 4G
#SBATCH -p main    # Partition
#SBATCH --qos main  # Quality of service
#SBATCH -o Salmon_sim_out-%J.txt  # STDOUT Out file
#SBATCH -e Salmon_sim_out-%J.txt  # Error file
#SBATCH -t 120                   # time in minuets



module load bio/salmon/1.3.0           # load the module


cd $SLURM_SUBMIT_DIR             # Moving to the test directory


#run salmon on the scripts
salmon quant -t transcripts.fasta --libType A -a sample_alignments.bam -o salmon_quan
t
```

Schedule the job with the following sbatch command.

sbatch myscript.sh

All the processed files will be generated in the same directory under salmon_quant.

Where to find help?

If you are stuck on some part or need help at any point, please contact OIT at the following address.

https://ua-app01.ua.edu/researchComputingPortal/public/oitHelp