TransDecoder on HPC cluster:

What is TransDecoder?

TransDecoder is a software package which identifies candidate coding regions within transcript sequences, such as those generated by de novo RNA-Seq transcript assembly using Trinity or constructed based on RNA-Seq alignments to the genome using Tophat and Cufflinks.

More documentation and tutorials can be found on the following links:

[Documentation](#)

[HPC Usuag](#)

Versions Available:

TransDecoder – 2.0.1

TransDecoder – 5.5.0

How to load a version of TransDecoder?

To load a version of TransDecoder on the HPC on current terminal session, use the following command:

module avail bio/transdecoder

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

module load bio/transdecoder/5.5.0

Verify by using this command:

module list

The loaded software and dependencies will be shown.

How to use TransDecoder on the cluster?

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

```
cd ~ && mkdir TransDecoder_test && cd TransDecoder_test    # Make a test directory on
home #folder


#Copy test data on server to the test dir

cp -r /share/apps/transdecoder/TransDecoder-TransDecoder-v5.5.0/sample_data/simple_tr
anscriptome_target/* ~/TransDecoder_test/


#unzip all files

gunzip *.gz



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

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/transdecoder/5.5.0


# start your commands here
```

```
# This method is ideal for short runs which produces runtime output
```

The Script:

```bash
#!/bin/bash
$SBATCH -J Jobname # Jobname
#SBATCH -n 1        # Nodes per task
#SBATCH –mem=10G     # Must need to allocate memory to run TransDecoder – at least 4G
#SBATCH -p main     # Partition
#SBATCH --qos main  # Quality of service
#SBATCH -o TransDecoder_sim_out-%J.txt  # STDOUT Out file
#SBATCH -e TransDecoder_sim_out-%J.txt  # Error file
#SBATCH -t 50                     # time in minuets



module load bio/transdecoder/5.5.0          # load the module


cd $SLURM_SUBMIT_DIR              # Moving to the test directory


# Use the transdecoder


TransDecoder.LongOrfs -t Trinity.fasta


TransDecoder.Predict -t Trinity.fasta


cdna_alignment_orf_to_genome_orf.pl  Trinity.fasta.transdecoder.gff3 genome_alignment
s.gmap.gff3 Trinity.fasta  > Trinity.fasta.transdecoder.genome.gff3
```

```
fasta_prot_checker.pl Trinity.fasta.transdecoder.pep
```

Schedule the job with the following sbatch command.

sbatch myscript.sh

All the processed files will be generated in the same directory as the script.

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