

BLAST on HPC cluster:

What is BLAST?

The NCBI Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.

The full documentation for the BLAST is found in the following links:

[Official Website](#)

[Documentation](#)

[Tutorial](#)

Versions Available:

- BLAST v2.3.0
- BLAST v2.9.0

How to load a version of BLAST?

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

```
module avail bio/blast/
```

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

```
module load bio/blast/2.9.0
```

Verify by using this command:

```
module list
```

Since this is a standalone software, it does not have dependencies. So, only BLAST will be seen listed.

Note: Important Programs in Blast Package

blastdb_aliastool

blastdbcheck

blastdbcmd

blast_formatter

blastn

blastp

blastx

convert2blastmask

deltablast

dustmasker

legacy_blast.pl

makeblastdb

makembindex

makeprofiledb

psiblast

rpsblast

rpstblastn

segmasker

tblastn

tblastx

update_blastdb.pl

windowmasker

How to use BLAST on the cluster?

There are two methods to run BLAST on the cluster.

The Interactive Way:

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/

# Start your commands here
blastx --help

#Follow with commands to execute
```

This method is ideal for a short job run which produces runtime output and to debug the codes.

The Script:

To run a slurm job, the user must prepare input files. For this example, get input files with,

```
#Download all the required files to the test folder
git clone https://github.com/enormandeu/ncbi_blast_tutorial.git

# Go to the cloned example test repository
cd ncbi_blast_tutorial
```

To submit this example job, or user's actual job, user must make bash script to submit to slurm scheduler,

```
# Make a script to schedule a job
touch script.sbatch
```

Use the following template for the script,

```
#!/bin/bash
#SBATCH -J test
#SBATCH -N 1
#SBATCH -c 8
#SBATCH --mem=16G
#SBATCH -p main
#SBATCH --qos main
#SBATCH -o BLAST-%J.txt
#SBATCH -e BLAST-%J.txt
#SBATCH -t 30

#Load the module

module load bio/blast

# Go to the submit directory
cd $SLURM_SUBMIT_DIR

#Run BLAST on the cluster

#Make a formatted database
makeblastdb -in reference.fasta -title reference -dbtype nucl -out
databases/reference

#Search sequence in the database
blastn -db databases/reference -query sequences.fasta -evaluate 1e-3 -
word_size 11 -outfmt 0 > sequences.reference

#####PARALLEL_EXECUTION#####
#####

# For parallel execution, use the following format | NOTE: GNU
```

```
PARALLEL MUST BE LOADED
time cat sequences.fasta | /share/apps/bin/parallel -k --block 1k --
recstart '>' --pipe 'blastn -db databases/reference -query - -evaluate
1e-3 -word_size 11 -outfmt 0' > sequences.reference
```

Schedule the job with the following sbatch command.

```
sbatch script.sbatch
```

All the processed files will be generated in the same directory as the sbatch script. User should see sequences.reference result file.

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>