Interproscan on HPC cluster:

What is InterProScan?

InterPro is a database which contains predictive information about proteins' function from several partner resources, giving an overview of the families that a protein belongs to and the domains and sites it contains.

To classify proteins in this way, InterPro uses predictive models, known as signatures, provided by several different databases (referred to as member databases) that make up the InterPro consortium.

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

InterPro Official Site

GitHub

Documentation

Versions Available:

- InterProScan v5.19

How to load a version of InterProScan?

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

```
module load bio/interproscan/5.19
```

Verify by using this command:

```
module list
```

Five dependencies should be loaded along with InterProScan itself.

Note:

To execute InterProScan, use the interproscan.sh command which will be added to the current PATH after loading the module.

How to use InterProScan on the cluster?

Create a script to run InterProScan on the cluster.

The Script:

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

```
#Copy the test input file from Interproscan installation directory
cp /share/apps/interproscan/5.19-58.0/test_proteins_new.fasta ~/test/

# Make a script to run Interproscan
touch script.sbatch
```

Note: ~/test refers to a test directory in the home folder. Feel free to use any other directory. Script should be in the same directory as in the input files.

Use the following template for the script,

```
#!/bin/bash
#SBATCH -J test
#SBATCH -p main
#SBATCH --qos main
#SBATCH -N 1
```

```
#SBATCH -c 16
#SBATCH --mem=16G
#SBATCH -e err.log
#SBATCH -o out.log

# Go to the script's directory
cd $SLURM_SUBMIT_DIR

#Load the module
module load bio/interproscan/5.19

#Run Interproscan           <file name
>          <workdir>    <temporary dir>
interproscan.sh -dp -i test_proteins_new.fasta -b ~/test/ -T /scratch/
```

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.

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