

What is BLAST?

The 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.

This guide to running BLAST provides instructions on how to run a standard BLAST test case on the HPC Wales systems.

Step 1 - Log in

The example used in this guide is configured to run on the Swansea Sandy Bridge cluster. Connect to *login.hpcwales.co.uk* with your HPC Wales user credentials using your preferred method (e.g. PuTTY from a Windows machine or ssh from any Linux terminal), then **ssh sw-sb-log-001** to connect to the Swansea system.

The steps below involve typing commands (**in bold font**) in the terminal window.

Step 2 - Load a BLAST module

A number of BLAST binary packages are available.

- List preinstalled BLAST versions:
module avail blast
- Load your preferred version (version 2.2.26 is used in this tutorial):
module load blast/2.2.26
- Confirm the loaded modules:
module list

Step 3 - Create a directory

Create a directory to hold any user data files. For this tutorial, a directory called BLAST should be created under your home directory:

```
cd ~  
mkdir BLAST
```

Step 4 - Obtain a test case

A test case for BLAST is provided with the installation at:

```
/app/genomics/blast/2.2.26/example/
```

Copy the SLURM job script:

```
cd ~/BLAST
cp -r /app/genomics/blast/2.2.26/example/run_blast.SLURM.q .
```

Note that the input files will be processed by the job using the /app directory identified above. The test case runs a search against the refseq_rna database as described at: http://www.ncbi.nlm.nih.gov/books/NBK52640/#chapter1.Setup_Steps_For_Legacy_blast.

Step 5 - Submit a job

BLAST is a serial application (i.e. it runs using just one core). The supplied test should run for less than one minute.

- Submit the job using:
`sbatch run_blast.SLURM.q`
- Check the job queue using:
`squeue`
- When completed, output can be found in a file called *blast.o.<JobID>* and errors, if any, can be found in *blast.e.<JobID>* (where <JobID> is the ID generated by the queuing system).
- If your job is successful, its output can be found in the output file called *refseq_blastall.out.<JobID>*
- Compare your job output with the reference output file

`/app/genomics/blast/2.2.26/example/refseq_blastall.out.`

Further info

Further information on BLAST can be found at the website <http://blast.ncbi.nlm.nih.gov/Blast.cgi> and a general introduction to BLAST can be found at <http://www.ncbi.nlm.nih.gov/books/NBK1734/>.