What is ClustalW?

ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. The home page for ClustalW is http://www.clustal.org/clustal2/ and the installed version 2 is described in the article available at http://bioinformatics.oxfordjournals.org/content/23/21/2947.long.

This guide to running ClustalW provides instructions on how to run a standard ClustalW 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 ClustalW module

- List preinstalled ClustalW versions:

  module avail clustalw

- Load your preferred version (version 2.1 is used in this tutorial):

  module load clustalw/2.1

- Confirm the loaded modules:

  module list

  Note: Dependent modules, such as the compiler environment, are loaded automatically.

Step 3 - Create a directory

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

  cd ~

  mkdir ClustalW

Step 4 - Obtain a test case

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

  /app/genomics/clustalw/2.1/example/

Copy the input files and job script:

```
cd ~/ClustalW
cp –rip /app/genomics/clustalw/2.1/example/* .
```

Step 5 - Submit a job

ClustalW is a serial application (i.e. it runs using just one core). The supplied test should run for only a few seconds.

- Submit the job using:

```
sbatch run_clustalw.SLURM.q
```

- Check the job queue using:

```
squeue
```

- When completed, output can be found in a file called ClustalW.o.<JobID> and errors, if any, can be found in ClustalW.e.<JobID> (where <JobID> is the ID generated by the queuing system).

- If your job is successful, its output can be found in the file being called ClustalW.log.<JobID>. Compare your job output with the reference output file ClustalW.log.

References

Further information on ClustalW can be found at the website: http://www.clustal.org/clustal2/

A description of the input to ClustalW is available at:
http://www.clustal.org/download/clustalw_help.txt.

Ref: HPCW-AG-15-011

hpcwales.co.uk

Ewrop & Chymru:
Buddsoddi yn eich dyfodol
Cronfa Datblygu Rhanbarthol Ewrop

Europe & Wales:
Investing in your future
European Regional Development Fund

ERDF

Llywodraeth Cymru
Welsh Government