



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 Supercomputing Wales systems.

Step 1 - Log in

The examples used in this guide is configured to run on the Cardiff Skylake *Hawk* cluster. Connect to *hawklogin.cf.ac.uk* with your Supercomputing user credentials using your preferred method (e.g. PuTTY from a Windows machine or ssh from any Linux terminal, thus

```
ssh -l your_username hawklogin.cf.ac.uk
```

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

Step 2 - Load a ClustalW module

In common with most of the application guides in this series, we assume at the outset that the module of choice would be selected from those originally available on HPC Wales. Thus a necessary first step would be to gain access to that module set by issuing the commands

```
module purge  
module load hpcw
```

Note that the “module purge” command is required to prevent any unintentional collision with pre-existing modules. Subsequently issuing the module command

```
module avail
```

will provide visibility of and access to the entire set of HPCW modules.

- 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, thus:

```
cd ~/ClustalW  
cp -r /app/genomics/clustalw/2.1/example/inputFiles .  
cp /app/genomics/clustalw/2.1/example/run_clustalw.SLURM.q .
```

Step 5 - Submit a job

ClustalW is a serial application (i.e. it runs using just one core) and is run on the htc partition of Hawk (though the SLURM directive "`#SBATCH -p htc`". 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.Hawk.o.<JobID> and errors, if any, can be found in ClustalW.Hawk.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.Hawk.log.<JobID>. Compare your job output with the reference output file ClustalW.Hawk.log.15512 in the directory /app/genomics/clustalw/2.1/example.

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.