Isak Sylvin
Swedish University of Agriculture (SLU)
2014-08-23

MAKER

Introduction

MAKER is an (almost) fully automated application for genome annotation. It easily integrates EST and protein homology data from public repositories to be used with a dozen of different prediction programs with different applications.

In this tutorial we will do two major exercises:

1) Dry run of MAKER with the prepackaged data
2) A full run of MAKER with fresh EST and protein data downloaded from NCBI

   If you have time, you may also try to quickly visualize the output.

The MAKER dry run

Log onto hpc.ilri.cgiar.org with mobaXterm

```
module load maker/2.28
cd ~
mkdir makerTutDry
cd makerTutDry
maker –CTL
ls
```

The maker –CTL runs the MAKER software to produce template versions of the configuration files. These must be changed so MAKER knows what type of analysis you want to use.

`nano maker_opts.ctl and change the following lines:`

```
genome=/export/apps/maker/2.28/data/dpp_contig.fasta
est=/export/apps/maker/2.28/data/dpp_est.fasta
protein=/export/apps/maker/2.28/data/dpp_protein.fasta
```

Save and exit.

`nano maker_boopts.ctl`

Make sure the settings look reasonable for the software in this file. If you want you may change the numbers here to get affect the quality of your results.

Save and exit.

Run:

```
maker maker_bopts.ctl maker_exe.ctl maker_opts.ctl
```

Once MAKER finishes running, first check the log file:

```
less ~/makerTutDry/dpp_contig.maker.output/dpp_contig_master_datastore_index.log
```

If everything worked as intended you should see:

```
contig-dpp-500-500      dpp_contig_datastore/05/1F/contig-dpp-500-500/  STARTED
contig-dpp-500-500      dpp_contig_datastore/05/1F/contig-dpp-500-500/  FINISHED
```

The final output of MAKER has been saved as a .gff file; view it with:
PLEASE NOTE THAT <TAB><TAB><TAB> MEANS YOU SHOULD PRESS TAB THRICE AT THAT POINT, NOT WRITE IT.

```
cd ~/makerTutDry/dpp_contig.maker.output/
less dpp_contig_datastore/<TAB><TAB><TAB>contig-dpp-500-500.gff
```

As you can see maker produces all output very nicely and intuitively; which can easily be exported into a genome browser for visualization.

Full run

In order to simulate a real run of MAKER we will download a small genome, EST and proteins for the species as well as EST from related organisms. For your convenience this data has been prepackaged and can be gotten with:

```
cd ~
wget http://hpc.ilri.cgiar.org/~isylvin/mycgen.zip
unzip mycgen.zip
ls
```

The files in question come from NCBI from the following sites (with some searching of course). It would be very easy for you to fetch the data yourselves in a normal case; but since we are running on HPC it's a bit trickier. We still provide the links so you see how easy it is to fetch supporting data. In case you're curious the data concerns the organism Mycoplasma Genitalium.

http://www.ncbi.nlm.nih.gov/nuccore/108885074?report=fasta Genome

http://www.ncbi.nlm.nih.gov/nucest Related EST

http://www.ncbi.nlm.nih.gov/nucest/?term=Mycoplasma%20genitalium Mycoplasma EST

http://www.ncbi.nlm.nih.gov/protein Protein

Since we are doing a new run we need new option files.

```
module load maker/2.28
cd ~
mkdir makerTutFull
cd makerTutFull
```

```
maker -CTL
ls
```

nano maker_opts.ctl and change the following lines where **XX** is replaced with the number next to your username. So if you're user20 the first line would be genome=/home/**user20**/mycgen/genMycGen.fasta:

```
genome=/home/userXX/mycgen/genMycGen.fasta
organism_type=prokaryotic
est=/home/userXX/mycgen/estMycGen.fasta
altest=/home/userXX/mycgen/altEstMycGen.fasta
protein=/home/userXX/mycgen/protMycGen.fasta
model_org=
prok_rm=1
cpus=3
```

Save and exit. Run:

```
interactive
maker maker_bopts.ctl maker_exe.ctl maker_opts.ctl
```

Analysis should take between five and ten minutes. Once MAKER finishes running, first check the log file:

```
less ~/makerTutFull/genMycGen.maker.output/genMycGen_master_datastore_index.log
```

If everything worked as intended you should see:

```
contig-dpp-500-500      dpp_contig_datastore/05/1F/contig-dpp-500-500/  STARTED
contig-dpp-500-500      dpp_contig_datastore/05/1F/contig-dpp-500-500/  FINISHED
```

The final output of MAKER has been saved as a .gff file; view it with:

```
cd ~/makerTutFull/genMycGen.maker.output/
less genMycGen_datastore/<TAB><TAB<TAB> gi%7C108885074%7Cref%7CNC_000908%2E2%7C.gff
```

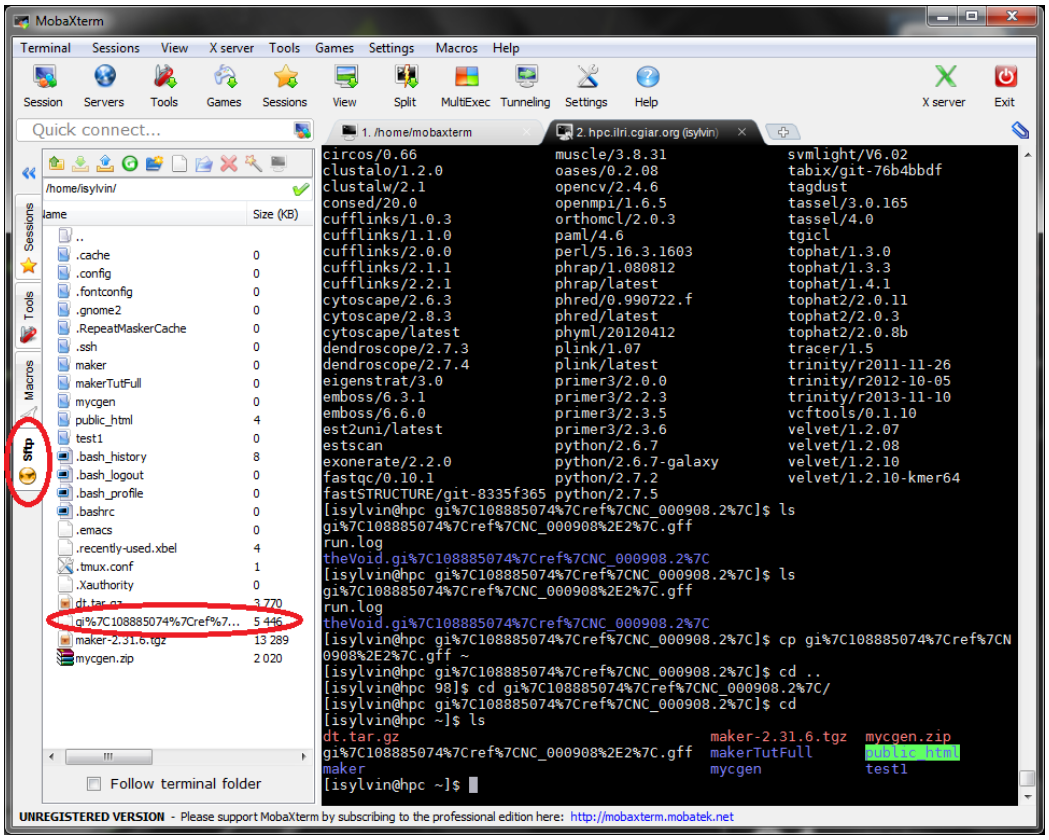
If You Have Time: Visualize your data

If you have time you might want to look at your now annotated data through a genome browser; the following can be done very quickly if you're interested.

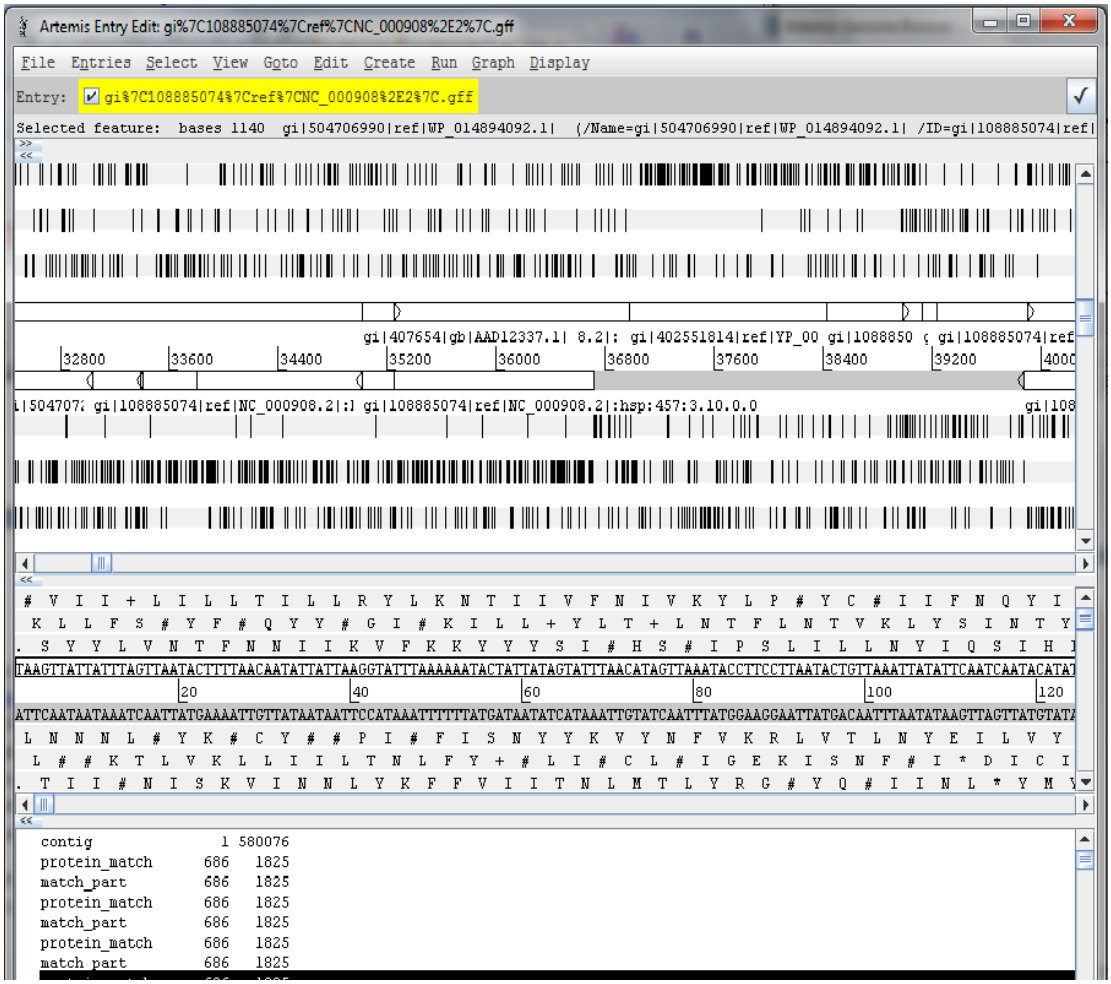
Do the following on one line (not two like in this document).

```
cp ~/makerTutFull/genMycGen.maker.output/genMycGen_datastore/<TAB><TAB<TAB>
gi%7C108885074%7Cref%7CNC_000908%2E2%7C.gff ~
```

You then have to copy the file to your desktop. Click on the SFTP tab in moba terminal; rightclick in the white box with all the files and select "Refresh current directory". Then drag the file named gi%7C108885074%7Cref%7CNC_000908%2E2%7C.gff from your white window and onto your desktop. **An image is attached to help you understand the process:**



Once on your desktop, download Artemis from https://www.sanger.ac.uk/resources/software/artemis/#download and start it (with a double click on windows, on Linux it needs to be started with the command line `java -jar artemis.jar` whilst in the download folder). Press OK; press File -> Open File Manager and select your file. It will most likely be located under
YOURUSERNAME`/Desktop/gi%7C108885074%7Cref%7CNC_000908%2E2%7C.gff`
If you get a warning just ignore it and press no.

If you've done everything correctly the genome browser will now produce the unannotated genome you downloaded with the genomes predicted by MAKER (see image).

Isak Sylvin
Swedish University of Agriculture (SLU)
2014-08-23