```
# Open terminal

# Log into your account
user<no.>@hpc01.icipe.org
password<no.>

# Determine whether you are in the home directory

pwd

cd

# Make a new directory

mkdir Assembly

# copy the fastq files you have into the new directory

mv ../Assembly.Lab/fastq_file_*.fastq ./Assembly

# Go into the new directory

cd Assembly

# Let's have a quick look at the files...

head fastq_file_1.fastq
head fastq_file_2.fastq

less fastq_file_1.fastq
less fastq_file_2.fastq

#lets count the lines in each file

wc -l  head fastq_file_1.fastq
wc -l  head fastq_file_2.fastq

# Now load all the software you require for this exercise
# If you don't know what's available you can first have a peek

module avail

# Now load Velvet

module load velvet/1.2.10
```

```
# Once velvet is loaded call velvet scripts to start work on the fastq
sequences
#First have a look at how to call Velvet

velveth -help

# now let's do the first step - running the hash (this creates a hash of
reads)

velveth my_assembly_directory 21 -shortPaired -fastq -separate
fastq_file_1.fastq fastq_file_2.fastq

# what files or directories were created

# Now that the first step is completed, lets do the next step - building the
graph

velvetg --help

velvetg my_assembly_directory –exp_cov auto –cov_cutoff auto

# The assembled sequence should be in the file contigs.fa, have a look at
this! What does it look like?

# At the moment you contigs are assembled but we don't know the order
# Now the next step would be to order you contigs

# We shall use abacas for that, so lets load it

module load abacas/1.3.1

# Abacas requires Mummer to run so let's also load Mummer

module load mummer/3.23

# Now let's run abacas, we shall require a reference sequence to order
against
# So let's use the provided sequence "reference_sequence.fa"

# Let's first see how to run abacas

 abacas -h

# Let's now run abacas
```

```
abacas -abcm -p nucmer -r reference_sequence.fa  -q
./my_assembly_directory/contigs.fa


# what files are produced?

ls -lrt

# Your ordered contig file is now called
"contigs.fa_reference_sequence.fa.MULTIFASTA.fa".

# You can convert this to a single fasta file using emboss

module load emboss/6.6.0

grep -v ">" contigs.fa_reference_sequence.fa.MULTIFASTA.fa | seqret -filter >
joined_file.dna

# Now that we have a single file we can do:

    1. Ab-initio gene prediction - i.e. predict genes in our assembled
sequence from scratch
    2. Annotation transfers using well annotated genomes as references

# Ab-initio gene prediction

module load   glimmer/3.02

# Now use glimmer

g3-iterated.csh joined_file.dna joined_file
glimmer3totab.pl joined_file.predict > joined_file.predict.tab

# Annotation transfers

mkdir Ref_embl
mv FM211187.embl ./Ref_embl

~/RATT/start.ratt.sh Ref_embl ~/my_assembly_director/single_fasta_file.dna
result Strain.Repetitive
```