

TAMU HPRC Short Courses – NGS Assembly – Noushin Ghaffari, PhD

Source/additional information:

<http://canu.readthedocs.io/en/latest/quick-start.html#quickstart>

=====
=====

Set up:

```
echo $SCRATCH
cd $SCRATCH
mkdir NGS_assembly_Oct18
mkdir NGS_assembly_Oct18/Long_Reads
mkdir NGS_assembly_Oct18/Long_Reads/PacBio_Assembly
mkdir NGS_assembly_Oct18/Long_Reads/PacBio_Data

cd NGS_assembly_Oct18/Long_Reads/PacBio_Data

curl -L -o pacbio.fastq http://gembox.cbcb.umd.edu/mhap/raw/
ecoli_p6_25x.filtered.fastq

head -n 16 pacbio.fastq
```

=====
=====

Running Canu on TAMU HPRC Ada System:

https://hprc.tamu.edu/wiki/Ada:NGS:Genome_Assembly#Canu

```
module spider Canu
module load Canu/1.7-intel-2017A-Perl-5.24.0
```

+++++

Running Canu:

```
head -n4 pacbio.fastq

cp /scratch/training/NGS_assembly/Scripts/Canu_All_PacBio.sh $SCRATCH/
NGS_assembly_Oct18/Long_Reads/PacBio_Assembly

cd $SCRATCH/NGS_assembly_Oct18/Long_Reads/PacBio_Assembly

cat Canu_All_PacBio.sh

bsub < Canu_All_PacBio.sh
```

bjobs

+++++

- Getting N50

cd /path_to_unitig_folder/unitigging

tgStoreDump -G *.gkpStore -T *.ctgStore 2 -sizes -s 4800000