Workflow for de novo assembly of genomic DNA on command line using Soap De Novo on Mason supercomputer:

In my account on Mason, I made folders as follows before loading the data:
mkdir NilsSeqs (to make a folder for all of the work I'm going to do on this project)
cd NilsSeqs (gets me into the folder I just made to create more folders and try to keep things organized from the get go)
mkdir Necator (to make folder for the raw data sequences)

To load data: On terminal in your home computer directory, use scp, choose file location then give destination. For example:

Lou-Ann-Bierwert-fcap:~ lbierwert$ scp (drag in file in mac or write out path on pc)
lbierwer@mason.indiana.edu:~/NilsSeqs/Necator (this tells my computer to upload data from a specific place on my computer to a folder that I just made in my account on the Mason supercomputer) If I did this right, I will be asked for my password to the supercomputer and then the upload will begin.

Cd to NilsSeqs

Mkdir NecatorOutput (make one more folder in your project folder for output)

Make a config file in nano: while in NilsSeqs folder type nano and hit return. This takes you to a simple word processing program that is easy to edit and write scripts here and then just send the program to this script to run.

In nano write:

```
max_rd_len=151
[LIB]
avg_ins=436
reverse_seq=0
asm_flags=3
rd_len_cutoff=143
rank=1
q1=Necator/NecatorAmericanis_S1_L001_R1_001.fastq
q2=Necator/NecatorAmericanis_S1_L001_R2_001.fastq
```

When done hit control o to "write out" which just means save. Save it as a good identifying name. I named it SoapConfigScript. I can go back to this general script and just change the fastq files to run this again.

The translation;

```
Max rd len = 151      (this is the max read length which I know is 151 because that was specified in the MiSeq run)
[LIB]                 (no idea I just know it has to be there)
avg_ins=436           (average insert size. This number came from the bioanalyzer)
reverse_seq=0         (will reverse all in q2 (second read) 0=R1 forward R2 reverse; 1= R2 forward, R1 reverse)
asm_flags=3           (will do both contigs and scaffolding)
rd_len_cutoff=143     (will cut all to this length – I usually do a FastQC in galaxy and trim there, that's how I get this number)
rank=1                (in which order reads are used while scaffolding)
q1=Necator/NecatorAmericanis_S1_L001_R1_001.fastq      (path to read 1 sequences)
q2=Necator/NecatorAmericanis_S1_L001_R1_001.fastq      (path to read 2 sequences, always after R1)
```

Now write out a qsub program and parameters in a second nano file in the same folder as the config. So make sure your in your folder, just cd NameOfFolder to be sure. Then type nano hit return and write:

```
#@job_type=serial
#@class=NORMAL
#@account_no=NONE
#PBS -m e -l vmem=48gb,walltime=48:00:00
```

```
#@notification=always
#@output=batch.$(cluster).out
#@error=batch.$(cluster).err
#@queue
cd /N/u/lbierwer/Mason/NilsSeqs
SOAPdenovo-31mer all -s SoapConfigScript -K 31 -R -o SoapDeNovo_Output
```