UF HPC Training
Working with NGS data
March 1, 2012

1. Get some data:
    a. Shared Data: Data Libraries: Training datasets: wine_yeast.100K.fq
2. NGS: QC and manipulation: **FASTQ Groomer**
    a. Input FASTQ quality scores type: Sanger
3. NGS: QC and manipulation:Fastqc: **Fastqc QC**
    a. Use the defaults or add a title for easier reference later
    b. Notice poor quality at ends of reads
4. NGS:QC and manipulation:**FASTQ Quality Timmer**
    a. Window size: 4
    b. Quality score: 30
    c. Rerun Fastqc QC on trimmed dataset
5. NGS: Assembly: **velveth**
    a. Hash Length: 29
    b. Click Add new Input Files
        i. File format: fastq
        ii. Dataset: Select your reads file
6. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velveth output
7. NGS: Assembly: **velvetg**
    a. Velvet Dataset: select the velveth output
    b. Set minimum contig length: yes
        i. Minimum contig length: 500
8. NGS: Mapping: **Map with Bowtie for Illumina**
    a. Will you select a reference genome from your history or use a built-in index?:
    b. Select S. cerevisiae (CYGD)

From the command line:

1. Login to submit
2. Go to you scratch space, make a directory called velvet_test and cd into it
3. Copy /project/bio/training/2012-03-01/velet.qsub to velvet_test
4. Edit the velvet.qsub file to have your e-mail.
5. qsub velvet.qsub