Bioinformatics and Sequencing Facility - Biology
Last Updated Wednesday, 01 April 2020 18:28

Description of the Facility

Background Overview
The Hunter College/CTBR Bioinformatics resources is located on the 4th floor of the Belfer Research Building at 69th Street and York Ave. The facility affords access to researchers and faculty, a high-performance computer cluster with a large range of bioinformatics software and data analysis pipelines. The facility provides cutting-edge bioinformatics technology for translational and basic research on health disparities. We also host a web-accessible bioinformatics platform based on Galaxy, (http://galaxy.hunter.cuny.edu:8080) to support genomic sequencing analysis.
Additionally, the facility offers Illumina Sequencing using the Illumina MiSeq sequencing platform and Nanopore sequencing using Oxford Nanopore MinIon sequencer. Both these instruments are capable of sequencing entire complement of DNA, or genome, of many animal, plant, and microbial species for basic biological and medical research. A detailed description of our services and available equipment is given below

### Services

- RNAseq and variation discovery
- small RNAs sequencing
- de novo bacterial genomes
- RNAseq Analysis
- Targeted amplicon sequencing
- Computational Capacity
- Scalable Storage

### Bioinformatics and Sequencing Resources and Equipment
Illumina MiSeq

MiSeq desktop sequencer:
Allows narrowly focused applications such as targeted gene sequencing, metagenetics, metagenomics, small genome and transcriptome sequencing, targeted gene expression, and amplicon sequencing.

Oxford Nanopore MinIon sequencer

MinIon portable sequencer: provides a rapid and portable, real-time sequencing platform that includes sequencing of full length transcripts with long reads, haplotype sequencing, metagenomic and 16S sequencing.

Agilent Technologies, 2100 Electrophoresis Bioanalyzer

The Agilent 2100 Bioanalyzer is a microfluidics-based platform that provides sizing, quantitation and quality control of DNA, RNA, proteins and cells on a single platform.

Two assay principles - electrophoresis and flow cytometry

Galaxy Web-accessible Bioinformatics Platform

We are running an installation of Galaxy, a web based platform for data intensive BioMedical research.

http://galaxy.hunter.cuny.edu:8080
Silicon Mechanics, HPC Cluster System

The high performance computing cluster provides 800 CPU cores, 3TB of high-speed RAM, a GPU Node for Visualizations, and a Docker node for virtualization.

- Redundant Head Node, 12 CPU Cores, 64 GB RAM
- 10 Compute Nodes, 20 CPU Cores each, 128 GB RAM
- 1 Medium Memory Node, 32 CPU Cores, 512 MB RAM
- 1 High Memory Node, 32 CPU Cores, 1 Terabyte RAM
- 1 GPU Node, K80, 2 CPU Hyper-Threaded / 128 GB RAM

ClusterStor 1500 solutions feature scale-out storage building blocks, the Lustre® parallel filesystem and a comprehensive management platform. The ClusterStor system provides TB's of ultra high speed data storage:

- 362TB of parallel storage
- 5GB/s throughput
- Seagate Enterprise Lustre
- Parallel based storage
The Belfer E-box provides storage for data backup and project archiving.
- 200TB of high availability storage
- 5GB/s throughput

Procedure For Submitting Jobs to the Cluster

Using SLURM, our job management platform. Virtual environments can be loaded using Conda our package manager. Once you activate your environment and you are sure your application is available, you can simply request `srun` or `sbatch`. Follow these steps to use SLURM:

1. Get a compute node assigned to you:
   
   ```
   [username@ctbr-cluster-hn1 env $] salloc
   salloc: Granted job allocation 48
   ```

   Salloc requests resources from the cluster. The number 48 in the example above is the allocation number provided to you. You can submit jobs directly to your allocation or run a batch script. Submitting a job using the command `srun` that will make use of the resources allocated. For this example we will use Bowtie2.

   ```
   srun bowtie2 -x Bowtie_2/hg38 -1 RNAseq_sample_data/adrenal_1.fastq -2 RNAseq_sample_data/adrenal_2.fastq -S alignment.sam
   ```

   `srun` = Directs the job to the compute node assigned by salloc.
   `bowtie2` = Main binary inside the virtual environment recently activated.
   `Bowtie_2/hg38` = This is the reference human genome
   `RNAseq_sample_data/adrenal_1.fastq` = First input strand
   `RNAseq_sample_data/adrenal_2.fastq` = Second input strand
   `alignment.sam` = The output file will be placed in the current directory

   A guide on using conda and SLURM can be obtained by clicking here.