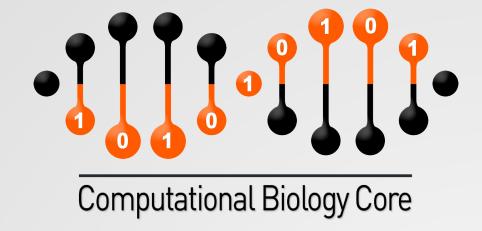
# Welcome! Bi-weekly Data Therapy

Vijender Singh Neranjan Perera Jill Wegrzyn





#### Online Support

Connect to the CBC team through



#bioinformatics\_help

uconn-cbc.slack.com

### **CBC Open House**

**Cell and Genome Sciences (UCH** Campus)

24th Feb 2017 9:30am - 4:30pm **Invited Speakers + GeneXPlain**  INSTITUTE FOR SYSTEMS GENOMICS

#### Computational Biology Core

Home People Hardware Software Databases Resources - FAQ Contact Us - Open House



#### Contact Information

#### Computational Biology Core

Institute for Systems Genomics Torrey Life Sciences Bldg., Rm. 375 75 N. Eagleville Rd, Unit-3043 Storrs, CT 06269-3043

JIII Wegrzyn, CBC Director/Assistant Professor Phone: (860) 486-8742

E-mail: jill.wegrzyn@uconn.edu

Vijender Singh, Lead Bioinformatic Scientist Phone: (860) 486-1324

E-mail: vijender.singh@uconn.edu

Neranjan Perera, Postdoctoral Researcher Phone: (860) 486-1324

E-mail: neranjan.perera@uconn.edu

Jeffrey W. Lary, Systems Administrator (BBC) Phone: (860) 486-5036

E-mail: jeffrey.lary@uconn.edu

Connect to the CBC team through

Stephen King, Systems Administrator (UCHC) Phone: (860) 679-7899

#bioinformatics\_help

uconn-cbc.slack.com

E-mail: stking@uchc.edu

Online Support

slack

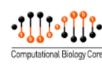
#### About Us

The Computational Biology Core (CBC) housed in the Institute for Systems Genomics at the University of Connecticut provides computational power and technical support to UConn researchers and affiliates. We collaborate with the Center for Genome Innovation (CGI) and Proteomics Core to provide full support to the research community.

#### Services Provided

- \* Experimental design consultation
- \* Computational analysis consultation
- \* Access to high performance computing systems
- \* Data storage and software installation service
- \* Biweekly group data therapy sessions
- \* Bioinformatics workshops and training





#### Upcoming ISG Events

MA: HCRAC Distinguished Investigator 4:00pm

All Events >

#### Bioinformatics Journal

#### Virtual exploration of early stage atherosclerosis

January 3, 2017

LAMSA: fast split read alignment with long approximate matches

September 25, 2016

Detecting subnetwork-level dynamic correlations

September 25, 2016

#### S Briefings in Bioinformatics

A network medicine approach to build a comprehensive atlas for the prognosis of human cancer

August 24, 2016

A comprehensive comparison of tools for differential ChIP-seq analysis

January 13, 2016

- · Center for Genome Innovation
- . UConn Biotech Center
- UCHC HPC Facility

#### Keep Informed!

**■Join** the **Bioinformatics Facility Mailing List** to learn about courses, seminars, and related discussion groups.

Follow us on Twitter!

#### Contact Us

#### Related Links

- Institute for Systems Genomics

- BECAT

#### Significant Bioinformatics Positions (ISCB)

An error has occurred, which probably means the feed is down. Try again later.



## CBC: Hardware

Two Clusters (Free Accounts)

BBC - Research and Teaching

HPC1 – Advanced Research

Xanadu – New Cluster Coming Online Soon!

- Access to more high memory machines



#### Hardware

To request an account on the cluster, please contact us. If you are planning a project with the CBC (or the CGI) and need to include a description of the computational resources, please use this text.

#### BBC Cluster (bbcsrv3.biotech.uconn.edu)

32-node Dell Linux cluster running Centos 6.3 and Rocks 6.1 x64.

#### Node configuration:

- 1: 16 x Quad-core 2.5 Ghz AMD 6370P processors with 512 GB RAM
- . 4: 4 x 8-core 2.00 GHz Intel Xeon processors with 64 GB RAM
- 17: 2 x Quad-core 2.53 GHz Intel Xeon processors with 32 GB RAM
- 10: 2 x Dual-core 2.8 GHz AMD processors with 16 GB RAM

Local storage capacity: Dell PowerVault MD1000 18TB RAID Array

#### UCHC HPC Cluster (hpc1-submit-int.cam.uchc.edu)

#### Node Configuration:

- 2,400 CPU cores with 5.8 TB RAM
- . 7,000 GPU cores with large CPU-only and hybrid compute clusters + OSG cluster

#### Virtualization Infrastructure:

 456 CPU cores, 1 TB RAM VMWare server and desktop virtualization hosts hosting 100+ Windows/Linux virtual machines with SSD high OPS performance cache tier

#### Datacenter Infrastructure:

- . UPS generator backed power with redundant cooling
- . 3x40 Gbe dark finer connection to off-site DR location

#### Network (100+ Gbe):

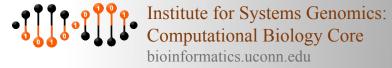
- Full non-oversubscribed 10/40 GbE datacenter network core layer
- BioScienceCT Research Network 100 GbE to CEN, Internet2, Storrs
- . New HPC Science DMZ low latency, 80 Gb-capable firewall

#### Storage (2+ PB):

394 TB EMC<sup>2</sup> Isilon and 832 TB Qumulo shared scale-out clusters along with 1,056 TB Amplistor on-premise cloud storage



Not our cluster (CERN) but here for inspiration



### **CBC:** Software

Home People Hardware Software Databases Resources - FAQ Contact Us - Open House

#### **Software**

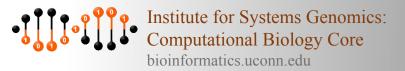
Software available on the clusters are freely available for use. Users must be comfortable in a <u>Unix environment</u> and understand how to properly submit jobs. Request additional information, accounts, and access here.

- 1. Annotation
- 2. Chip Seq
- 3. Genome Assembly
- 4. Metagenomics
- 5. Molecular Structure
- 6. Multiple Sequence Aligners
- 7. Phylogenetics
- 8. Population Genetics
- 9. Proteomics
- 10. RAD-Seq
- 11. Repeat Analysis
- 12. RNA-Seq
- 13. Sequence Clustering
- 14. Sequence File Manip./Quality Control
- 15. Sequence Submission
- 16. Short Read Aligners
- 17. Single Cell Genomics
- 18. Statistics
- 19. Transcriptome Assembly
- 20. Variation Detection
- 21. Visualization Tools

### Two Clusters (Free Accounts)

BBC and HPC1 with core Bioinformatics Applications

- > 300 Applications Supported on BBC
- Open Source
- Variety of Genetics, Genomics, and Proteomics Analysis



### **CBC:** Tutorials

### **Custom Guides:**

Interacting with the Cluster Analysis and Viz with R! RNA-Seq Genome Assembly

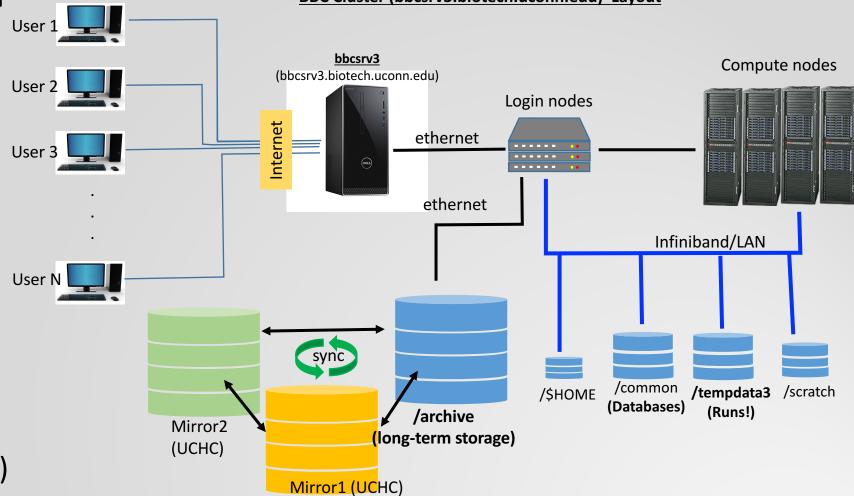
More on the way!



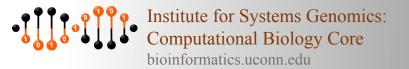
#### **Tutorials Tutorial** Last Updated Description Server Access UConn Health Cluster (PBS) June 2015 Understanding the UConn Health Cluster BBC Cluster (SGE) June 2015 Understanding the BBC Cluster UNIX and R Unix Basics November Introduction to Command Line Operations 2013 VIM Introduction to VIM UNIX Editor December 2013 **Basic Bioinformatic Exercises** Unix Examples November 2013 Introduction to R November Basic Analysis and Plots 2013 **RNA-Seq Guides** EDGE-pro tutorial (with Listeria reference genome) Prokaryote RNA-Seq (EDGE-pro/DESeq2) July 2015 Model Plant RNA-Seq (STAR/DESeq2) July 2015 RNASeg tutorial (with Glycine max reference genome) Non-Model Plant RNA-Seq August 2015 RNA-Seq tutorial (with Picea rubens reference (Bowtie2/eXpress/DESeq2 transcriptome) Human RNA-Seq (no replicates) (STAR/DESeq2) March 2015 Introduction to RNASeq Model Insect RNASeq (Web-based Galaxy) RNASeq tutorial (with Drosophila reference genome) July 2015 **Genome Assembly** Genome Size Estimation Tutorial January 2017 Genome Size Estimation Tutorial Bacterial Genome Assembly Tutorial September Genome Assembly tutorial 2015

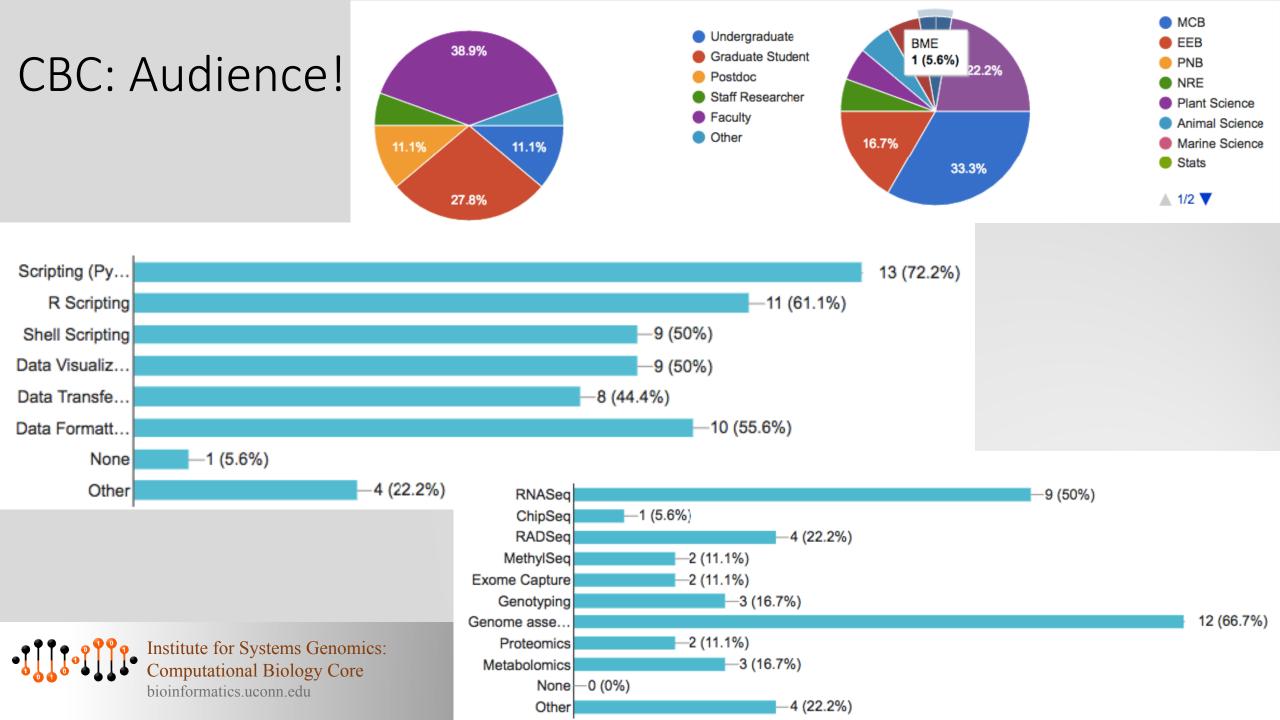
CBC: BBC Cluster

BBC Cluster (bbcsrv3.biotech.uconn.edu) Layout



Submitting Jobs!
Where to Run and Store?
Long-term storage (/archive)





## CBC: Organize

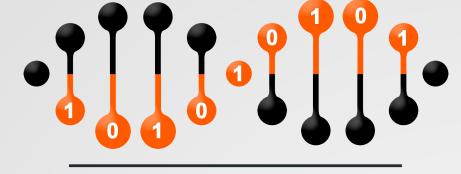
Invited Speakers
Research Talks
Directed Q&A
Code Exercises
Mini Tutorials

## **Stay Connected!**





Bioinformatics-I@Listserv.uconn.edu



Computational Biology Core

