



## ABSTRACT

The Bioinformatics Facility of the Biotechnology Resource Center (BRC) established and maintains the BioHPC Computing Laboratory, a cloud computing resource configured for life sciences research. The BioHPC Lab provides investigators with direct access to a wide range of bioinformatics data analysis software tools. Users can access the Lab's computers either remotely or on-site in the core facility. The Lab has servers with configurations designed for different type of data analysis, including large memory servers for *de novo* sequence assembly analysis. Many of the most commonly used data analysis software for next generation sequencing are on the Lab's computers. Training on software analysis tools is available. Lab workstations are reserved using an online scheduler. Research groups can also have their own workstations hosted in the Lab, maintained by the facility staff but accessible only to the host group's members. The Lab is available to the university community and to outside investigators. The BioHPC Computing Laboratory is a key resource for a diverse array of research projects.

## OVERVIEW

**General:** The BRC Bioinformatics Facility's BioHPC Computing Laboratory is a cloud computational resource configured for biologists. The Lab is targeted for biologists who want Linux or Windows operating system, and to do bioinformatics data analysis themselves. The Lab provides investigators with direct access to a wide range of bioinformatics data analysis software tools on appropriate hardware platforms. The available hardware ranges from small 8-core 16 GB RAM servers to large 96-core 512 GB RAM machines and 64 core 1024 TB RAM machines.

**Access:** Workstations are reserved using an online calendar-based scheduler; only registered Lab users can make reservations. All workstations are accessible remotely with ssh or VNC; selected workstations are accessible on-site.

**Fees:** The BioHPC Computing Lab is a fee based service system. Users need to purchase computing hours to make reservations. Users also purchase storage if their needs are in excess of the Lab's free storage allocation.

**Storage:** The BioHPC Computing Lab has over 473TB of network storage. Each registered active user gets 200 GB storage space; users can purchase additional storage.

**Training:** The Bioinformatics Facility offers workshops introducing biologists to the Linux computing environment. Users can also talk to us during our office hours (Monday and Thursday, 1:00-3:00 pm, appointment required) for consultation and training.

**Administration:** The BioHPC Computing Laboratory is maintained by the Bioinformatics Facility of the Biotechnology Resource Center, which is part of the Cornell University Institute of Biotechnology.

**Open to all:** The resources and services of the Bioinformatics Facility are open to all investigators at Cornell University and Cornell-affiliated institutions. The facility also provides services to external investigators at both academic institutions and commercial enterprises. In association with the NY State Center for Advanced Technology (CAT), NY State companies receive a discount on facility services.

## RESOURCES

Physical and Virtual Computing Laboratory



Remotely accessible through Internet 425 Rhodes Hall  
24/7 access with passcode lock

## Software

282 software titles available. New software can be added by request. Development tools are also available for deploying custom programs.

Name	Version	Name	Version	Name	Version
Htgenomics	1.2.0	HMMER	3.1b1	QUMA	1.0.0
454 gAssembler					
gMapper	2.8	HOPSNOT	4	R	3.2.5
a5	20140401	HTSeq	0.6.1	RACA	0.9.1
ABYSS	1.9.0	HyPhy	g63c18d	RAPTR-SV	0.0.6
AdapterRemoval	2.1.1	Assembler	1.3.2	RAML	8.1.17
AdminTools	1.1	BRD	3.11	Ray	2.1.1
AdminSuite	1.23	IDBA-ID	1.1.1	REAPR	1.0.18
Alker	1.03	IGV	2.3.12	RepeatMasker	4.0.5
AlignSeq	1.11	IMAG	2.4	RepeatMasker	1.0.8
ALLMAPS	20150710	impuz2	2.3.0	RFMix	4.0.5
ALLPATHS-LG	40148	InteProScan	5.17.56.0	RNASEM	1.2
AMON	3.1.0	InteProScan	5.17.56.0	RNEM	12.29
analysis	0.8.4	jellyfish	2.2.3	RSeQC	2.6.1
ANOSD	0.911	JobShop	4.1	RI	0.99.901
Annovar	20130823	julia	0.3.5	subre	1
Atlas Link	2050709	jpegtran	4.0.6	Sigmap/GW	0.1
ATLAS Capfill	2.2	ksutils	0.424	StenoSeq	1.3
ATSAS	2.5.0.2	Kent src utilities	11/9/2009	Satsuma	3.1
Augustus	2.2.1	klmer	0.4	scika-beam	20160828
bamtools	2.5.0	kMILin	0	scycle	12/18/2013
BayesScan	2.1	LDAK	4.9	SeqPrep	12/17/2013
Bitmap	35.66	bc64	8/13/2014	agrus	1.94a
Beagle	3.3.2	LocustZoom	1.3	SILAPET	v2.790
Beagle4	r1399	LUCY	1.20p	shore	0.9.3
bedops	2.4.15	LUCY2	2.19p	SIREMBP	3.2
BEDTools	2.22.1	LUMPY	0.2.9	SIREMBP	2.2.3
BigSnpSites	1.01	MACS	1.4.2	stacks	12/18/2013
bioBandman	0.0.125	MAQS2	2.1.0	SupraP	4.1
Bioconductor	3	MAFFT	7.273	skover	0.1.127
BioByt	1.6.922	MAKER	2.32	SMET Analysis	2.20
BioPython	1.6	MAQ	0.7.1	unakenake	3.6.1
Biodreda	1.5.5	MaskCA	3.1.0	sup	1.0beta.17
Biomart	0.7.7	Maave	2.4.0	SNVlyse	20140701
blat	9/24/2014	mcortext	5/30/2014	SOAP2	soap2.21
BLAST	v2.3.0	megahit	1.0.3.7	SOAPdenovo	1.05
	v1.2.1	DR		SOAPdenovo	
blast2go	Mar2016	MEGAN	5.10.0	Tron	1.03
BLAT	34	MEME Suite	4.10.1.4	SOAPdenovo2	2.4
Bowtie	1.1.2	MEMBERS	v1.563	SIPules	3.7.1
Bowtie2	2.2.8	MetaMap	1.7.8	SRA_Poolkit	2.5.7
BUSCO	1.2	MetaVelvet	1.2.02	stacks	1.37
BWA	0.7.13	MetaVelvet-SL	1	stampy	1.0.22
causal	1.1	mica	4.0.2	STAR	2.5
CAP3	12/21/2007	mildRept2	0.0.5	starmodels	0.6.1
CBSI RNAseq		MIRN (mmapy)	0.4.9	StratTet	1.0.3
cd-hit	4.6.1	McMapper	2	STRUCTURE	2.3.4
CGMAP	2.5	MkTest	2	SURPI	5-Jan-15
Circos	0.67.0	MMSEQ	1.0.0.beta2	sun	3/19/2012
Circoscape	4.0.5	MtBayes	3.2.2	SVDetect	0.8b
Chrom-Omega		multFAST	3.1.6	svmapsum	2014_08_15
CLUSTALW	2.1	mid	20150708	tabix	1.3
Cluster	1.52	MSMC	9/15/2015	Tandem Repeats Finder (TRF)	4.04
		MSR-CA		utils	
		Genome			
CNVator	0.3	Assembler	1.9.4	TASSEL 3	3.0.158
CronMap	0.2.3	metat	0.32	TASSEL 4	4.3.6
Cuffdiff	2.2.1	MetaMap	20140810	TASSEL 5	
cutadapt	1.8.1	mgpy	vi2.2	icoflee	11
DocuSeq	1.7.0	MUMmer	3.23	TEItoolkit	2.06
DocuSeq	0.4.3	NCBI	3.4.31	EMBM4	2.1.1
delly	0.6.5	Nemo	2.2.0	TopHat	2.1.1
DETONATE	1.1	Neofusion	7.0.1	transRate	1.1.2
diamond	0.7.5	NEURON	7.3	TransDecoder	2.1.0
Disovar	52488	new_fusion	2016_06_02	transrate	1.0.1
Disovar de novo	52488	NeoFusionMap	0.4.11	trinity	1.12
DREG	20160428	NGSdatax		trinityomatic	2.3.6
dk	14/706	nglDht	v1.0.2	Trinity	0.2.0
dnvutils	re9822	nglF	v1.0.0	Trinotate	3.0.0
		USC_Kent			
cooper	0.5.0	ngTools	20140114	utils	Sep. 22 2015
		Variant Effect Predictor			
ERGENSOFT	5.0.1	NGSUtils			7.7
EMBOSS	64.0.0	ngsutils	2.08.03	veclidkit	0.2.4
exabayes	1.4.1	NovoalignCS	1.02.03	veeCoker	1.1
exonerate	2.2.0	Novoalign	0.2.20	veff	3/13/2015
eXpress	1.15.2	NovoAlign	Jun-11	veffools	0.1.14
FALCON	0.4.0	OrthoMcl	V1.4	Velvet	1.2.10
FastML	20150324	PAML	4.8	VismatRNA	2.2.4
fastq_species_detect					
fastq	1	panDas	0.12.0	VirulIndex 2	2
FastQC	0.11.5	panseq	2.6	vsarech	1.1.3
FastTree	2.1.7	Panseq		wgs-assembler	8.3ec2
PASTX	0.0.13	PASTEC	1	(Geneweb)	2.4.1
fastSTRUCTURE	2.06	pbtools	0.8.0	yaha	0.1.82
flex	12.10	PhyML	12.8.14		
Flexible Adapter Remover	2.15	PhyML	15.8.24		
fastp	v1.0.2	PeakSplitter	1		
GATK	3.5	PEAR	0.9.6		
GCTA	1.25.2	PeakSV	Jun-11		
GEMMA	0.94	pbtools	0.8.0		
genef	1.4.4	PHAST	1.3		
genef	4.3	PIVLP	3.096		
GenoMark	latest	PhysCSF	2012028		
GenoMarker		prech	20120412		
		Genome STRIP	1.03.619	PhyML	20131031
		GenomeMapper	0.4.4	Picard	2.1.1
		GenomeStudio	2011.1	PinDel	0.2.4
		germline	1.5.1	PIQ	1.3
				1.09_beta3.3	
		GMAP/GSNAP	5/15/2014	plink	0
		GNI/Compass	1.3.0	popln	0.3
		GNI/parallel	20150422	prismc	0.20.2
		Gubbins	20150324	prodigal	2.6.2
				propriser/Clustal	
		HapCompass	0.8.1	s	0
		hapR	1.2.1	pyRAD	3.0.5
		HaplotypeMerger	20120810	QIMM	1.9.1
		HaplotypeMerger2	20151124	Quake	0.3.5
			5-Jan-13	QUAST	4

## Software Support



- Comprehensive list of bioinformatics software is maintained and updated by core facility staff.
- Users guides and basic instructions for each software package are available on the core facility web site.
- New applications added based on user requests. If you need a program that is not installed, contact us!
- We will install any applications that benefit a sizable group of users and that do not require extensive development effort.
- If a program requires significant effort to install and/or maintain and is not essential for a sizable group of users, we can still install it, but we will charge for the effort.

## Hardware



Multiple tiers of hardware resources are available configured for different data analysis applications, including systems with very big RAM and a large number of cores.

- Computing: interactive (accessible on site and remote)**
- 3 x 4 core (1 CPU), 24 GB RAM, 4 TB storage, NVIDIA Quadro NVS 295 Graphics Card, 24 inch LCD monitor. (Linux)
  - 1 x 4 core (1 CPU), 16 GB RAM, 10 TB local storage, NVIDIA Quadro FX570 Graphics Card, 24 inch LCD monitor. (Windows)

- Computing: general**
- 32 x Dell PowerEdge M600 blade server, 8 core (2 Intel CPUs), 16 GB RAM, 1 TB HD local storage. (Linux)

- Computing: medium memory**
- 16 x 128 GB RAM machines; each 12 core (2 Intel CPUs), 128 GB RAM, 1 TB SSD fast storage, 4 TB regular SATA storage. (Linux)

- Computing: large memory**
- 3 x Gen2 512 GB machines; 64 core (4 AMD CPUs), 512 GB RAM, 1 TB SSD fast storage, 9 TB regular SATA storage (4 x 3 TB RAIDs). (Linux)

- 8 x Gen1 512 GB machines; 64 core (4 AMD CPUs), 512 GB RAM, 12 TB regular SATA storage (6 x 3 TB RAIDs). (Linux)
- 1 x Gen3 512 GB machines; 64 core (4 Intel E5 CPUs), 1024 GB RAM, 16 TB regular SATA storage (6 x 4 TB RAIDs). (Linux)
- 2 x Gen2 512 GB machines; 96 core (4 Intel E7 CPUs), 512 GB RAM, 1 TB SSD fast storage 12 TB regular SATA storage (6 x 4 TB RAID10). (Linux)

- Login nodes**
- 2 x 6 core (1 Intel CPU), 64 GB RAM, 1 TB regular storage. (Linux)

- Storage (Lustre & Gluster)**
- 2 x 4 core (1 CPU), 24 GB RAM with 12 disk rack enclosure (4 TB disks) connected to LSI 9285-8e RAID controller configured at RAID6
  - 4 x 6 core (1 CPU), 64 GB RAM, 24 x 4 TB disks configured at RAID6
  - 5 x 8 core (1 CPU), 64 GB RAM, 24 x 4 TB disks configured at JBOD
  - 2 x 8 core (2 CPU), 256 GB RAM, 6 x 16 TB SSD disks in RAID10
  - 2 x 8 core (2 CPU), 64 GB RAM, 2 x 16 TB disks in RAID1

## Future



- More applications! We are adding applications regularly based on demand.
- Extension to the cloud. We plan to add option to run BioHPC Lab servers in the cloud using the same configuration as our local servers, i.e. with access to our software and storage.

## Consultation, Workshops and Training



Examples of recent workshops:

- Linux for biologists
- Perl for biologists
- Variant Calling
- Transcriptome assembly
- RNA-Seq Data Analysis
- Gene Function Annotation
- Genome Assembly
- ChIP-Seq Data Analysis and Motif Calling

□ The BioHPC Computing Lab is used for bioinformatics workshops and for several university courses.

□ One-to-one training is provided to Lab users during weekly office hours.

□ Consultation on project design and data analysis available upon request. The BRC Bioinformatics Facility and Genomics Facility provide coordinated joint consultation services for next generation sequencing projects. This includes consultation on the selection and optimal application of software tools for data analysis.

□ Bioinformatics software analysis tools and training available through the facility.

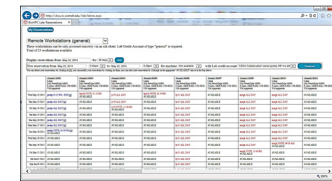
□ Educational workshops and training on data analysis available through the facility.

□ Coordinated project design consultation and data analysis support available with the BRC genomics, genomic diversity proteomics and mass spectrometry, imaging, bio-IT, and advanced technology assessment core facilities.

## Access

Calendar Based "Computing Hour" Reservation System

- Machines can be reserved using an online calendar scheduler
- Users can share reservations
  - the account owner can add additional users
- Reservations can be cancelled anytime online
  - only the part used is charged



## Contact Information



Bioinformatics Facility  
Jaroslav Pillardy, Director  
Qi Sun, Co-Director  
BRC\_bioinformatics@cornell.edu  
www.biotech.cornell.edu/brc/bioinformatics



For questions about the Biotechnology Resource Center please contact George Grills at BRC\_biotech@cornell.edu

www.biotech.cornell.edu

