

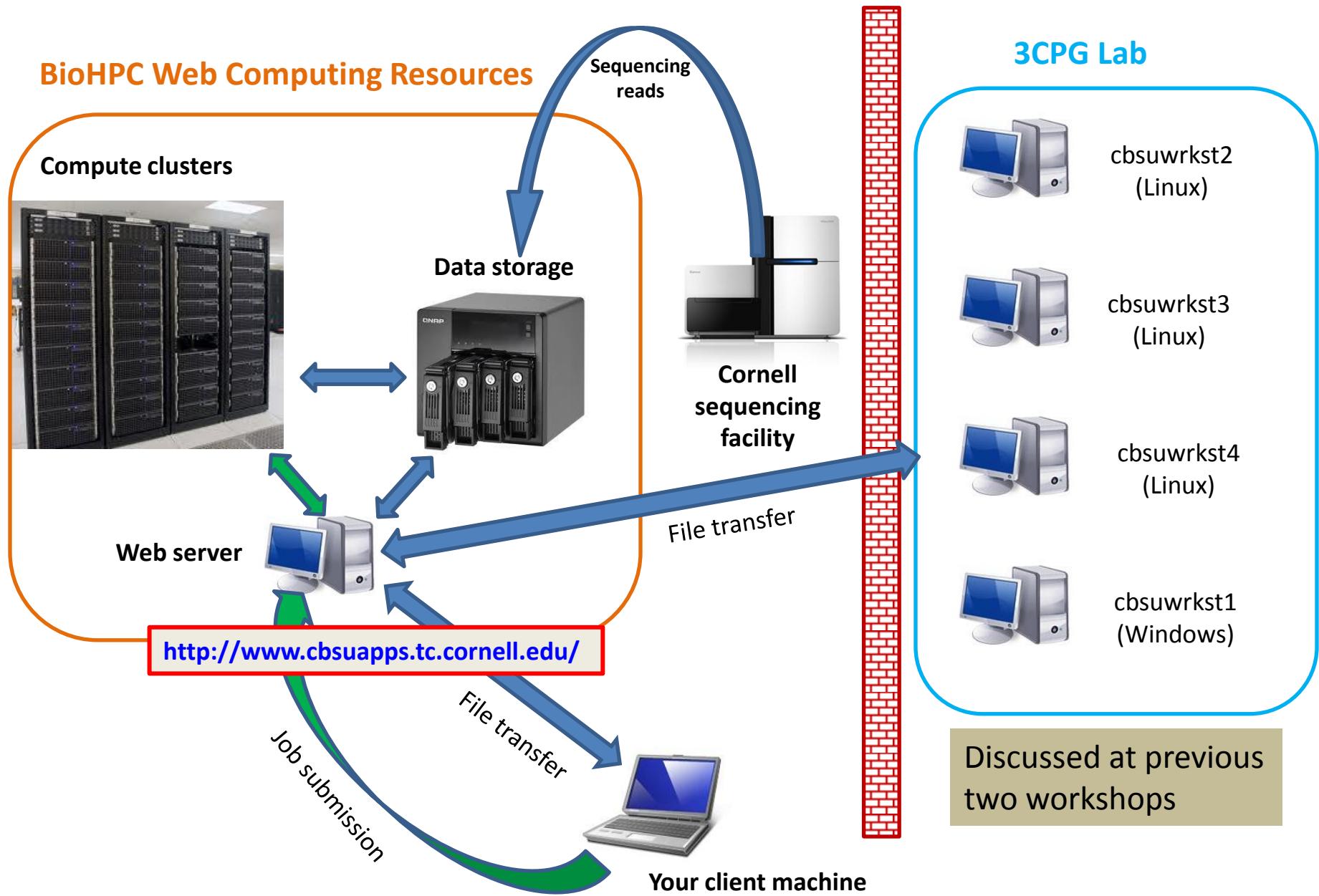
# **BioHPC Web Computing Resources at CBSU**

3CPG workshop

Robert Bukowski  
Computational Biology Service Unit

[http://cbsu.tc.cornell.edu/lab/doc/BioHPC\\_web\\_tutorial.pdf](http://cbsu.tc.cornell.edu/lab/doc/BioHPC_web_tutorial.pdf)

# BioHPC infrastructure at CBSU



# BioHPC Web Computing Resources at CBSU

- Have been around for 10 years, with Next-Gen support started recently
- **Compute clusters**
  - Currently about 1000 CPU cores
  - 250 cores on machines suitable for Next-Gen data analysis (the exact number will depend on demand)
  - A large memory (64 GB) machine
  - Looking to upgrade the aging hardware
- **Data store**
  - Combined 15 TB of storage
  - For calculations only, NOT to be treated as permanent
    - File retention policy: kept for 30 days since the date it was deposited (in practice: much longer)
- **BioHPC Suite**
  - Collection of 40+ open source computational biology applications, including 7 Next-Gen data analysis programs (so far)
  - **BioHPC Web Interface**
    - **Submission pages:** for submitting applications to BioHPC compute clusters
    - **Data Manager:** interface to the data store
    - **Pipeline Manager:** a tool for constructing simple analysis pipelines (beta version) - see tutorial at [http://cbsuapps.tc.cornell.edu/doc/Pipelines\\_Manual.pdf](http://cbsuapps.tc.cornell.edu/doc/Pipelines_Manual.pdf)

# BioHPC Web Interface

The screenshot shows a Microsoft Internet Explorer window displaying the CBSU Web Applications interface. A red circle highlights the address bar, which shows the URL <http://cbsuapps.tc.cornell.edu/index.aspx>. A red arrow points from the left towards the 'Login' link in the user status bar.

**CBSU Web Applications - Internet Explorer provided by Dell**

User: guest | [Login](#) | [Apps Home](#) | [FAQ](#) |

**APPLICATIONS**  
(click on a category below to access programs)

Show all Hide all

- [Next-Gen Tools](#)
- [Sequence analysis](#)
- [Sequence alignment](#)
- [Population genetics](#)
- [Protein structure](#)
- [MSR Biomedical](#)
- [Other](#)
- [Links](#)

## Computational Biology Service Unit BioHPC Web Computing Resources

[\(compute nodes status\)](#)

### Welcome to CBSU Web Computing Interface

As a part of our mission we are designing easy-to-use web interfaces for various computational biology tools. This application suite and interface is a product of long development of an user friendly system providing access to high performance computing for computational biology. The development started in 2001 as a CBSU project and over the years changed into [Computational Biology Applications Suite for High Performance Computing \(BioHPC Suite\)](#). This suite is available for download along with the interface source code and will run/interface with any Microsoft-based cluster. You can find more details on the [BioHPC Suite home page](#). The BioHPC project was supported in 2006-2008 under [Microsoft High-Performance Computing Institutes](#) (CBSU was one of them) and currently is supported by Microsoft Research.

The applications are run on CBSU computer clusters as well as on Microsoft Athena cluster. For more information about our hardware resources can be found [here](#). Microsoft gave us access to their 64 node cluster in their headquarters running MS Windows HPC Server 2008, the cluster's name is Athena and it is being accessed via *HPC profile / JSDL protocol*. It is an example of how well BioHPC can integrate geographically dispersed computing resources.

The usage of various applications in this BioHPC installation is summarized on [BioHPC@CBSU statistics page](#).

Some of the tools are freely available for all, some, due to a very high computational demand, are available to registered users only. Current user id, together with user-specific functions, is displayed in the bar above.

We spent a lot of time and effort developing this interface, and we hope it is a useful resource. You can help us in this effort by acknowledging our contribution to your work and research. Please acknowledge us in all publications and presentation of work that used our resources using the following [text](#). Please send us information about your publications and presentations to the address below. Better yet - take our [survey](#). Thanks!

# Logging in to BioHPC Web Interface

- Account required to use Next-Gen applications
- Account separate from 3CPG lab account
- Your e-mail address is your login ID
- Many of you already have an account on BioHPC web interface
  - anyone who used the system before
  - anyone who submitted a sample for sequencing to Cornell facility

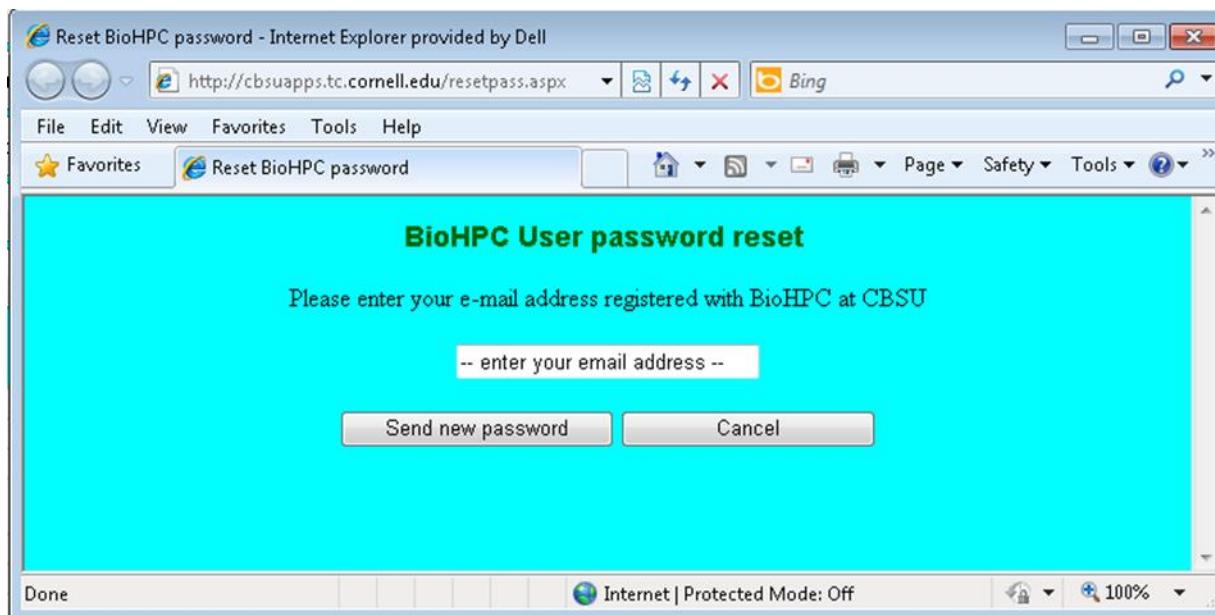
The screenshot shows a web browser window with a light blue header bar containing the URL "l=index.aspx". Below the header is a navigation bar with "Previous" and "Next" buttons, a search icon, and an "Options" dropdown menu. The main content area has a teal background and features a title "CBSU User login" in green. Below the title is a login form with two rows. The first row contains a "Email" label and a text input field containing "bukowski@cac.cornell.edu". The second row contains a "Password" label and a text input field filled with seven black dots. At the bottom of the form is a "Login" button.

Email	bukowski@cac.cornell.edu
Password	•••••••

Login

# Logging in to BioHPC Web Interface

- ❑ To obtain/re-set password – try  
<http://cbsuapps.tc.cornell.edu/resetpass.aspx>
- ❑ If your e-mail address is not recognized – contact us at  
<http://cbsuapps.tc.cornell.edu/contactus.aspx> to register



# Next Gen applications in BioHPC Web

The screenshot shows a Microsoft Internet Explorer window displaying the CBSU Web Applications homepage. The title bar reads "CBSU Web Applications - Internet Explorer provided by Dell" and the address bar shows "http://cbsuapps.tc.cornell.edu/index.aspx". The menu bar includes File, Edit, View, Favorites, Tools, and Help. A toolbar below the menu bar has a Favorites icon, a CBSU Web Applications icon, a search field with placeholder "-o", and buttons for Previous, Next, Options, and a pencil icon.

The main content area features the CBSU logo (CB and SU with DNA helixes) and the text "Computational Biology Service Unit" and "BioHPC Web Computing Resources". Below this is a link to "(compute nodes status)". The user information "User: bukowski@cac.cornell.edu" is shown along with links to Apps Home, FAQ, Logout, Change password, and My jobs.

A sidebar on the left is titled "APPLICATIONS" with the sub-instruction "(click on a category below to access programs)". It contains two buttons: "Show all" and "Hide all". A red arrow points to the "Next-Gen Tools" link. A second red arrow points to the "Next-Gen Help" link. The sidebar also lists other tools: Data Manager, Bowtie, Bowtie-build, BWA, BWA index, Cufflinks, FASTX, RNASeq, SamTools, and TopHat.

The main content area has a large heading "Welcome to CBSU Web Computing Interface". Below it is a detailed text block about the BioHPC suite's history and current status. Another text block discusses the hardware resources and software stack. A third text block provides a summary of application usage statistics. A final note at the bottom states that some tools are freely available while others require registration.

**1<sup>st</sup> step to getting help**

# BioHPC Web Resources FAQ

FAQ - Internet Explorer provided by Dell

http://cbsuapps.tc.cornell.edu/faq.aspx

File Edit View Favorites Tools Help

Favorites FAQ

Find: -o Previous Next Options

 Computational Biology Service Unit  
BioHPC Web Computing Resources  
(compute nodes status)

User: bukowski@cac.cornell.edu | [Apps Home](#) | [FAQ](#) | [Logout](#) | [Change password](#) | [My jobs](#) |

**APPLICATIONS**  
(click on a category below to access programs)

Show all Hide all

**Next-Gen Tools**

**Sequence analysis**

**Sequence alignment**

**Population genetics**

**Protein structure**

**MSR Biomedical**

**Other**

**Links**

## Frequently Asked Questions

**General**

[What is BioHPC?](#)  
[How do I run my analysis?](#)  
[How do I register with CBSU?](#)  
[What are the benefits of being a registered user?](#)  
[How do I log in to the CBSU web page?](#)  
[How do I unpack a "gzipped tarball" on Windows?](#)  
[My job finished prematurely. Why? Is there a way to run it longer?](#)  
[What is "cluster maintenance" and how does it affect my jobs?](#)  
[I want to submit to Athena, but my jobs get re-routed to other clusters. Why?](#)  
[Your server says there something wrong with my e-mail address, but I'm sure it is correct](#)  
[While submitting a job, I get a message that I have other active jobs in the system](#)  
[Which e-mail addresses can I use to submit a job?](#)

**Next Generation Sequencing data analysis**

[How to run Next-Gen applications offered by BioHPC?](#) ←

**BEAST**

[Can I extend the timeout of my BEAST run?](#)

**IM / IMa**

[Can I extend the timeout of my IMa run?](#)

# Example: BWA job submission

BWA @ BioHPC - Internet Explorer provided by Dell  
http://cbsuapps.tc.cornell.edu/bwa.aspx

File Edit View Favorites Tools Help

Favorites BWA @ BioHPC

Find: Previous Next Options

**APPLICATIONS**  
(click on a category below to access programs)

Show all Hide all

**Next-Gen Tools**

- Next-Gen Help**
- Data Manager**
- Bowtie**
- Bowtie-build**
- BWA**
- BWA index**
- Cufflinks**
- FASTX**
- RNASeq**
- SamTools**
- TopHat**

**Sequence analysis**

- Sequence alignment**
- Population genetics**

**Protein structure**

**MSR Biomedical**

**Other**

**Links**

## BWA @ BioHPC

Version 0.5.8c (r1536)  
Please send comments to [biohpc@cornell.edu](mailto:biohpc@cornell.edu).

**BWA** (Burrows-Wheeler Aligner) is a fast light-weighted tool that aligns relatively short sequences (queries) to a sequence database (target), such as the human reference genome. This web page uses **one** of the two different algorithms implemented in BWA, designed for short queries up to ~200bp with low error rate (<3%). It does gapped global alignment w.r.t. queries, supports paired-end reads, and is one of the fastest short read alignment algorithms to date while also visiting suboptimal hits. The second algorithm, BWA-SW, designed for long reads with more errors is **not** yet supported by this submission page.

To use BWA to align your reads to a reference database, the BWA-formatted **index** of this database must be available. A database is selected using the **reference genome index file** selector (see below). If the list does not contain the index of the database you need, you will have to first create this index from the FASTA file using the **BWA index** application (click "Next Gen Tools", on the left, then "BWA index") and then return to this page.

Calculations will be carried out on the BioHPC compute cluster at [CBSU](#). You will receive e-mail notifications when the job is submitted, when it starts, and when it is finished. Output will be available via links embedded in the notification e-mails. For more information about this program and BioHPC interface in general, please visit our [Frequently Asked Questions](#) page.

Please acknowledge us in all publications and presentation of work that used our resources using the following [text](#).

**Specify reference genome index file:**

-- select file(s) --

Don't see the index you need? Use the **BWA index** application to make one.

**Read file(s) type:**  unpaired reads  paired-end reads

**Specify read file:**

-- select file(s) --

Input is selected from among the files present in BioHPC data store.

Dropdowns show:

- Only files with proper format
- Only files you have access to
- Don't see your file? We'll show how to upload it to BioHPC store

# Example: BWA job submission, cont.

**Required program options:**

general options

Name of the output file (extension ".gz" or ".bam" will be appended)  
 Create output in BAM format

aln options

**[ -n ]**. Maximum edit distance if the value is INT, or the fraction of missing alignments given 2% uniform base error rate if FLOAT. In the latter case, the maximum edit distance is automatically chosen for different read lengths.  
 **[ -o ]**. Maximum number of gap opens  
 **[ -e ]**. Maximum number of gap extensions, -1 for k-difference mode (disallowing long gaps)

samse options

**[ -n ]**. Maximum number of alignments to output in the XA tag for reads paired properly. If a read has more than INT hits, the XA tag will not be written.

**Advanced program options (if unsure, leave blank):**

aln options:

samse/sampe options:

Register output for future use within BioHPC  
Enter short description of output to be registered:  
This is a test alignment file produced at a workshop

Cluster: Auto ( [Show timeout info](#) )

**Submit** **Reset**

What happens if NOT checked?

- You will still be able to download the output file(s), BUT
- These files will not be seen by jobs you may want to run next.

# Job submission confirmation

Submitting your job ... please wait  
registering job info ...

```
query type unpaired  
selected read files: 450  
number of read files selected: 1output file name my_test_alignment
```

```
email bukowski@cac.cornell.edu  
* cluster => cbsumic1b001  
* user => bukowski@cac.cornell.edu  
* job id => 208030  
... done  
creating directories ...  
script ccp.bat done  
script file 2 done  
... done  
InputData successfully serialized and saved  
submitting job to cluster ...
```

Job ID

```
* cluster job id => 8  
... done  
finalizing db entries done  
Your BWA job my_test_alignment (208030) has been SUBMITTED
```

You will receive another email once your job is finished.

You may follow program's progress by viewing [here](#)

Timeout information and the current job status can be found [here](#)

If you want to CANCEL the job, please click [here](#)

The alignment file can be downloaded via http [here](#) and (upon job completion) via ftp [here](#)

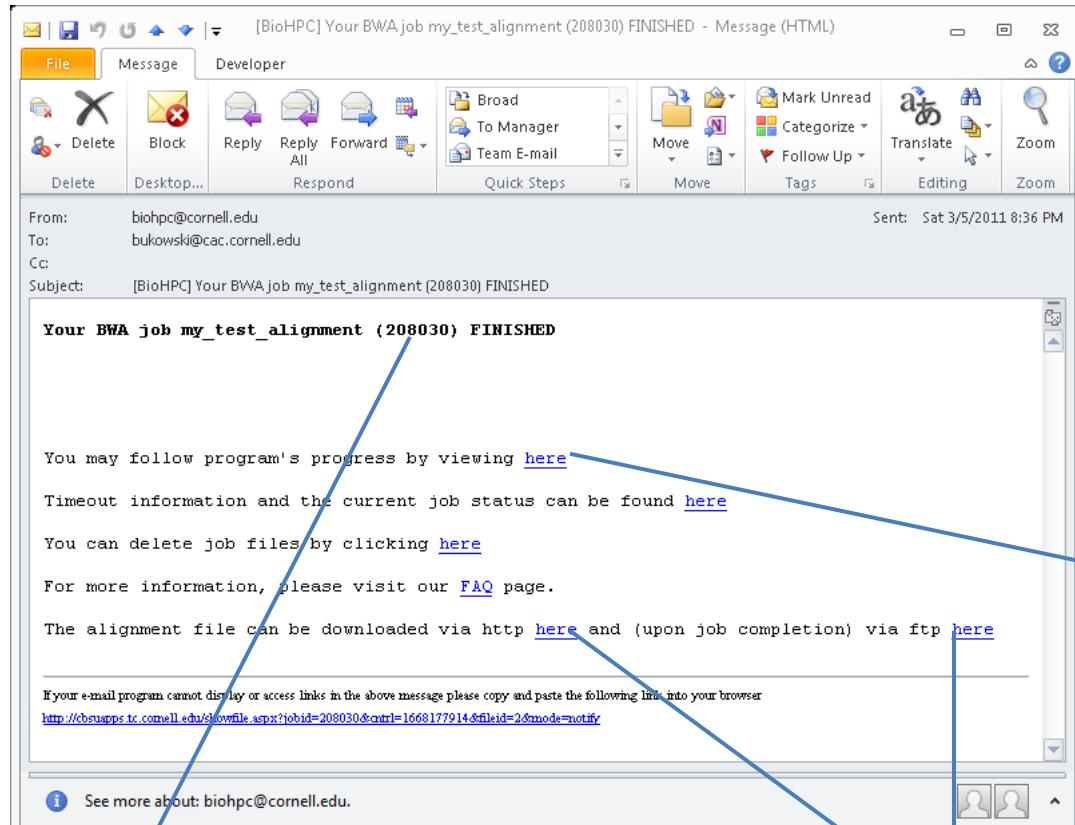
- When you see this page, you are DONE. You can close the browser or continue working (maybe submit another job).
- Notifications about the job and links to results will be **e-mailed** to you.

## What is happening behind the scenes:

- Job is entered in a queue on a compute cluster.
- Job scheduler on the cluster will decide when to start the job.
- Wait time (from submission to start) depends on the load.

Job submission confirmation – you will receive an e-mail with this information.

# Job notification e-mails



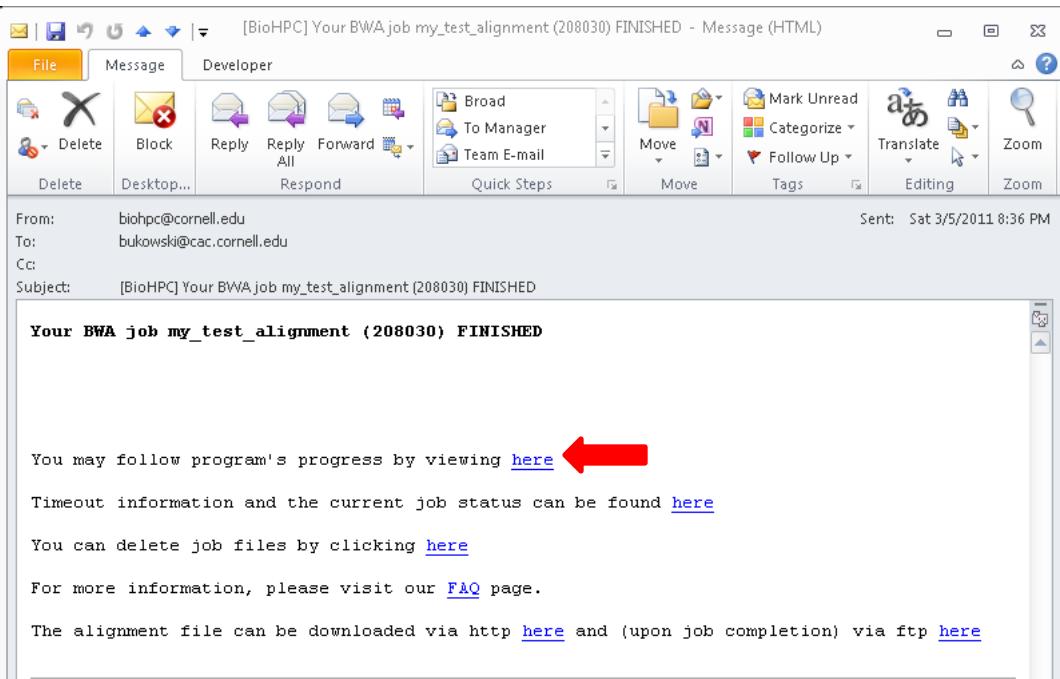
Job ID

This is where output can be seen and/or downloaded (the exact message depends on application)

Sent when

- Job is submitted
- Job starts (may be a while after submission, depending on system load)
- Job finishes

This link allows you to monitor progress while the job is running.



If your e-mail program cannot display or access links in the above message please copy and paste the URL into a web browser:  
<http://cbsuapps.tc.cornell.edu/showfile.aspx?jobid=208030&cntrl=1668177914&fileid=2>

See more about: biohpc@cornell.edu.

```
Copying job data to local scratch
Fetching BWA index files
Fetching read file(s)
Running BWA with command: /scratch/serloopp/CBSU/JOB$/208030/bwa aln -n 0.04 -o 1 -e -1 ../../indexes/Ecoli_NC_009800_BWAindex_e_coli_10000snp.fq 1>
[bwa_aln] 17bp reads: max_diff = 2
[bwa_aln] 39bp reads: max_diff = 3
[bwa_aln] 64bp reads: max_diff = 4
[bwa_aln] 93bp reads: max_diff = 5
[bwa_aln] 124bp reads: max_diff = 6
[bwa_aln] 157bp reads: max_diff = 7
[bwa_aln] 190bp reads: max_diff = 8
[bwa_aln] 225bp reads: max_diff = 9
[bwa_aln_core] calculate SA coordinate... 0.36 sec
[bwa_aln_core] write to the disk... 0.00 sec
[bwa_aln_core] 10000 sequences have been processed.
Running BWA with command: /scratch/serloopp/CBSU/JOB$/208030/bwa samse -n 3 ../../indexes/Ecoli_NC_009800_BWAindex_e_coli_10000snp.fq 1>m
[bwa_aln_core] convert to sequence coordinate... 0.02 sec
[bwa_aln_core] refine gapped alignments... 0.01 sec
[bwa_aln_core] print alignments... 0.02 sec
[bwa_aln_core] 10000 sequences have been processed.
Removing files no longer needed
Converting SAM to BAM: /scratch/serloopp/CBSU/JOB$/208030/samtools view -bS -o my_test_alignment.bam my_test_alignment 2>>my_test_alignment.log
[samopen] SAM header is present: 1 sequences.
```

[BioHPC] Your BWA job my\_test\_alignment (208030) FINISHED - Message (HTML)

**File** **Message** **Developer**

Delete Desktop... Respond Quick Steps Move

From: biohpc@cornell.edu  
To: bukowski@cac.cornell.edu  
Cc:  
Subject: [BioHPC] Your BWA job my\_test\_alignment (208030) FINISHED

**Your BWA job my\_test\_alignment (208030) FINISHED**

You may follow program's progress by viewing [here](#)

Timeout information and the current job status can be found [here](#)

You can delete job files by clicking [here](#)

For more information, please visit our [FAQ](#) page.

The alignment file can be downloaded via http [here](#) and (upon job comp

If your e-mail program cannot display or access links in the above message please copy and paste the following link into your browser  
<http://cbsum1c1b001/cbsujob/showfile.aspx?jobid=208030&curl=1668177914&fileid=2&mode=notify>

 See more about: biohpc@cornell.edu.

cbsujob - Internet Explorer provided by Dell

File Edit View Favorites Tools Help

Favorites cbsujob

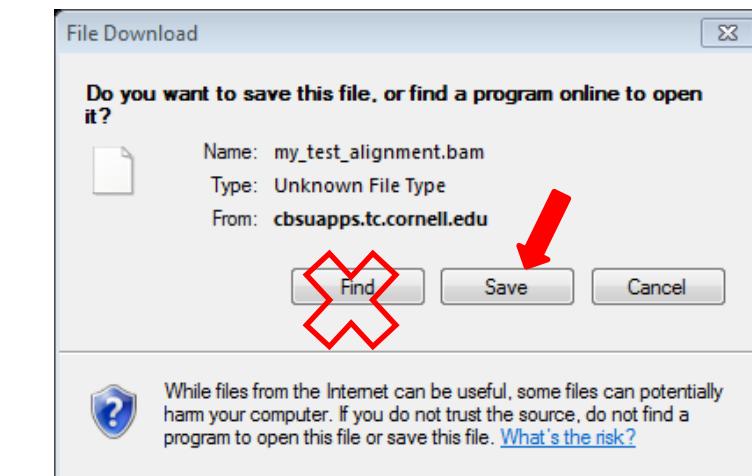
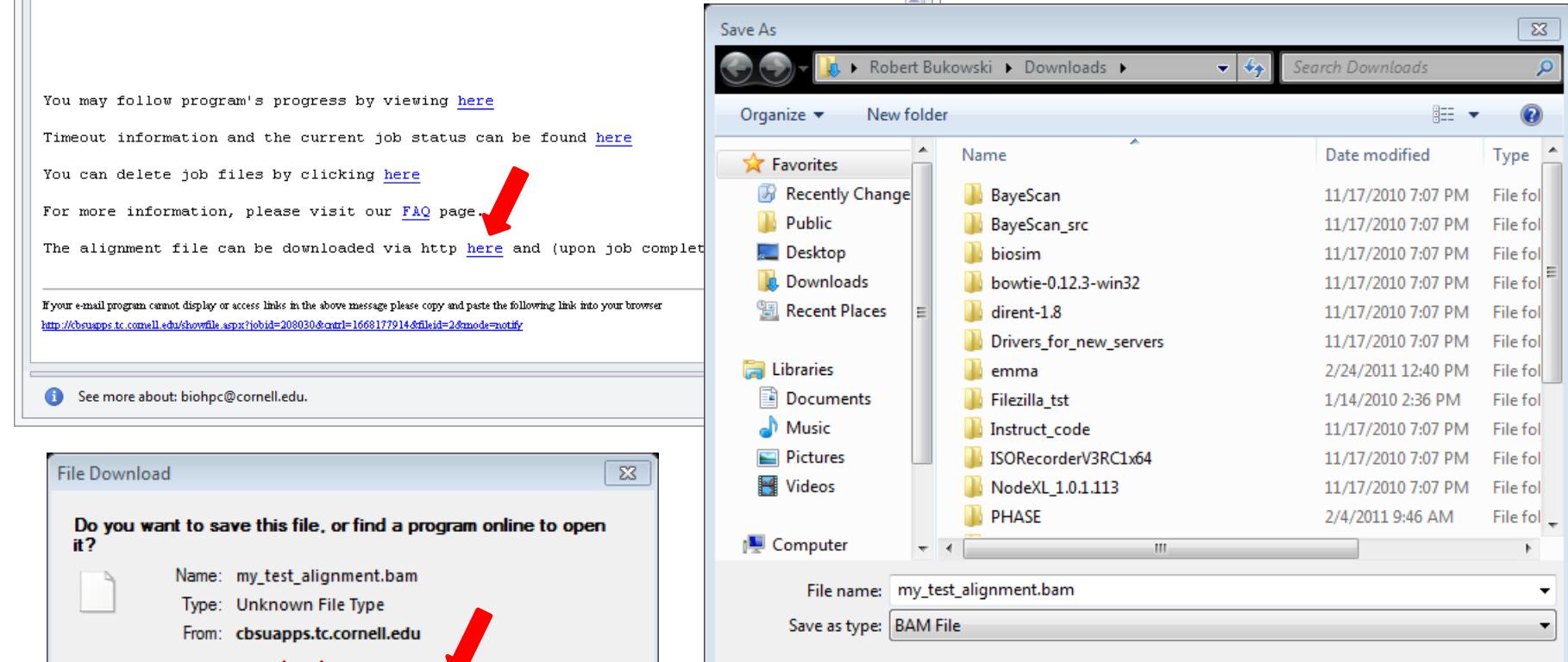
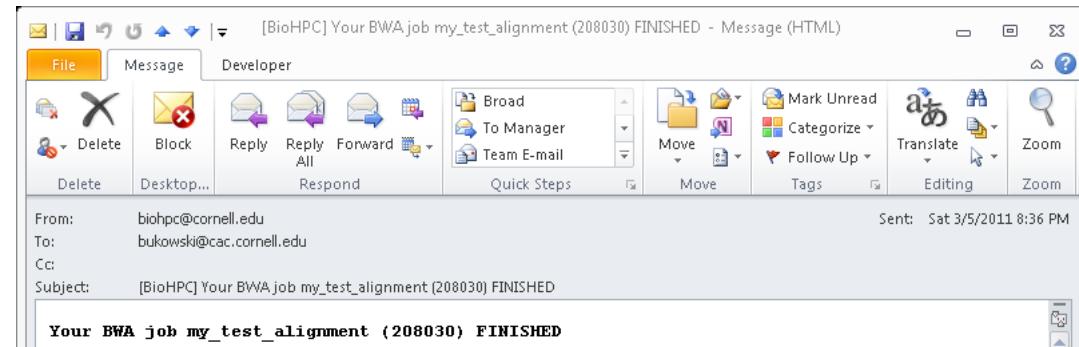
**JOB STATISTICS:**

Job ID:	208030
Job name:	my_test_alignment
Application:	BWA
Cluster:	cbsum1c1b001
CPUs requested:	1
Contact e-mail:	bukowski@cac.cornell.edu
Timeout:	2880 minutes
Submitted:	3/5/2011 8:31:06 PM
Started:	3/5/2011 8:31:06 PM
Ending:	3/5/2011 8:35:42 PM
Status:	FINISHED CORRECTLY

OK

Internet | Protected Mode: On

100%



**Note:** if the result BAM file will be used only with BioHPC Web applications, you don't have to download it.

# What if my job fails?

- ❑ When your job fails, you will receive the notification e-mail about it.
- ❑ Even failed jobs usually produce some output (log files, for example), which often contains clues about the reason for the failure.
- ❑ Download and examine all output files (check **all** links in the notification e-mail) opening them in a text editor
  - Often the error messages you'll find in those files clearly point to formatting problems or incorrect command line options. Look for words like "error", "failed", etc.
  - Fix these problems and re-submit the job
- ❑ If you cannot determine the reason for failure, contact us at  
<http://cbsuapps.tc.cornell.edu/contactus.aspx> (please specify the **Job ID** – you can find it in any notification e-mail sent about this job).

# Advantage of BWA @ BioHPC Web

Recall how many steps it takes to obtain a BAM alignment file on a Linux workstation  
(from Qi Sun's workshop 3/2/2011):

```
gunzip s_1_sequence.txt.gz
bwa aln -n 2 indexes/Ecoli_NC009800_BWAind s_1_sequence.txt > s_1.sai
bwa samse -n 5 maize.fa s_1.sai s_1_sequence.txt > s_1.sam
samtools view -bS -o my_test_alignment.bam s_1.sam
```

The BioHPC Web interface to BWA will take care of all this for you in one click.  
No need to reserve time on Lab workstations  
No need to deal with Linux

# CBSU RNA-Seq @ BioHPC

another simple interface to a complex pipeline

RNASeq @ BioHPC - Internet Explorer provided by Dell  
http://cbsuapps.tc.cornell.edu/RNASeq.aspx

File Edit View Favorites Tools Help

Favorites RNASeq @ BioHPC

Find: -o Previous Next Options

Computational Biology Service Unit  
BioHPC Web Computing Resources  
(compute nodes status)

User: bukowski@cac.cornell.edu | Apps Home | FAQ | Logout | Change password | My jobs |

APPLICATIONS  
(click on a category below to access programs)

Show all Hide all

Next-Gen Tools

- Next-Gen Help
- Data Manager
- Bowtie
- Bowtie-build
- BWA
- BWA index
- Cufflinks
- FASTX
- RNASeq
- SamTools
- TopHat

Sequence analysis

- Sequence alignment

Population genetics

Protein structure

MSR Biomedical

Other

Links

CBSU RNA-Seq @ BioHPC  
(version 1.0)  
Please send comments to [biohpc@cornell.edu](mailto:biohpc@cornell.edu).

RNASeq PIPELINE DESCRIPTION:

The transcriptomics pipeline consists of the following steps:  
(1) Parsing the annotation to extract exons, exon-junctions, transcripts and genes  
(2) Indexing the genome  
(3) Aligning reads against the genome (can be somewhat time-consuming!)  
(4) Aligning the reads that do not align to the genome against the exon-junctions  
(5) Calculating the coverage of exons, transcripts and genes

The parsing of the annotation and the indexing of the genome have already been run, just choose which genome annotation you are running your sequence against.

OUTPUT:

Gene coverage file contains the following 5 columns:

- > gene\_name
- > name of the highest-rpm transcript-isform for the gene
- > length of the selected transcript
- > rpm of the selected transcript
- > weighted coverage of the selected transcript

References

Nat Genet. 2010 Oct 31.

[The developmental dynamics of the maize leaf transcriptome.](#)

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR, Reidel EJ, Turgeon R, Liu P, Sun Q, Nelson T, Brutnell TP.  
[1] Boyce Thompson Institute, Cornell University, Ithaca, New York, USA. [2] These authors contributed equally to this work.

PMID: 21037569 [PubMed - as supplied by publisher]

Statistics

- BioHPC Home
- CBSU Home
- CBSU ftp server
- CBSU SeqDB
- CTC Windows
- Bioinformatics Applications
- DISTRUCT
- T-REX (T-RFLP manager)
- Next-Gen@BioHPC
- CBSU Survey
- Read Survey (adm)
- Reset Password
- F A Q
- Contact Us

Select aligner:  BWA  Bowtie

Select genome annotation: Maize AGPV2

Select sequence file:  
-- select file(s) --

Register output files ( sam files only) for future use within BioHPC  
Enter short description of output file to be registered:

Version 1 Rev 429  
(2011/02/18 11:52:31)

Options:

prefix  Prefix for intermediate and output file names.  
 Uniquely-mapped reads only.

Cluster: Auto [\( Show timeout info \)](#)

Submit Reset

# So, do I still need Linux?

Currently  
implemented:



Yes, if

- Need a tool which is not (yet) available on BioHPC Web
- Need custom scripting throughout the project
- Need interactivity and experimentation with parameters
- Need to run a graphical application (e.g., iAssembler, IGV)

Parts of the project (e.g., alignment) may be completed using BioHPC Web Resources, other parts – on Linux.

# Files on BioHPC Data Store

Before any Next-Gen application can be run via BioHPC web interface, all the input files must be present and catalogued on the BioHPC data store.

Files automatically deposited at BioHPC data store:

- Illumina sequencing data files from Cornell sequencing facility
- Files produced by jobs run through BioHPC web interface (if requested by checking the “Register output for future use within BioHPC” checkbox).

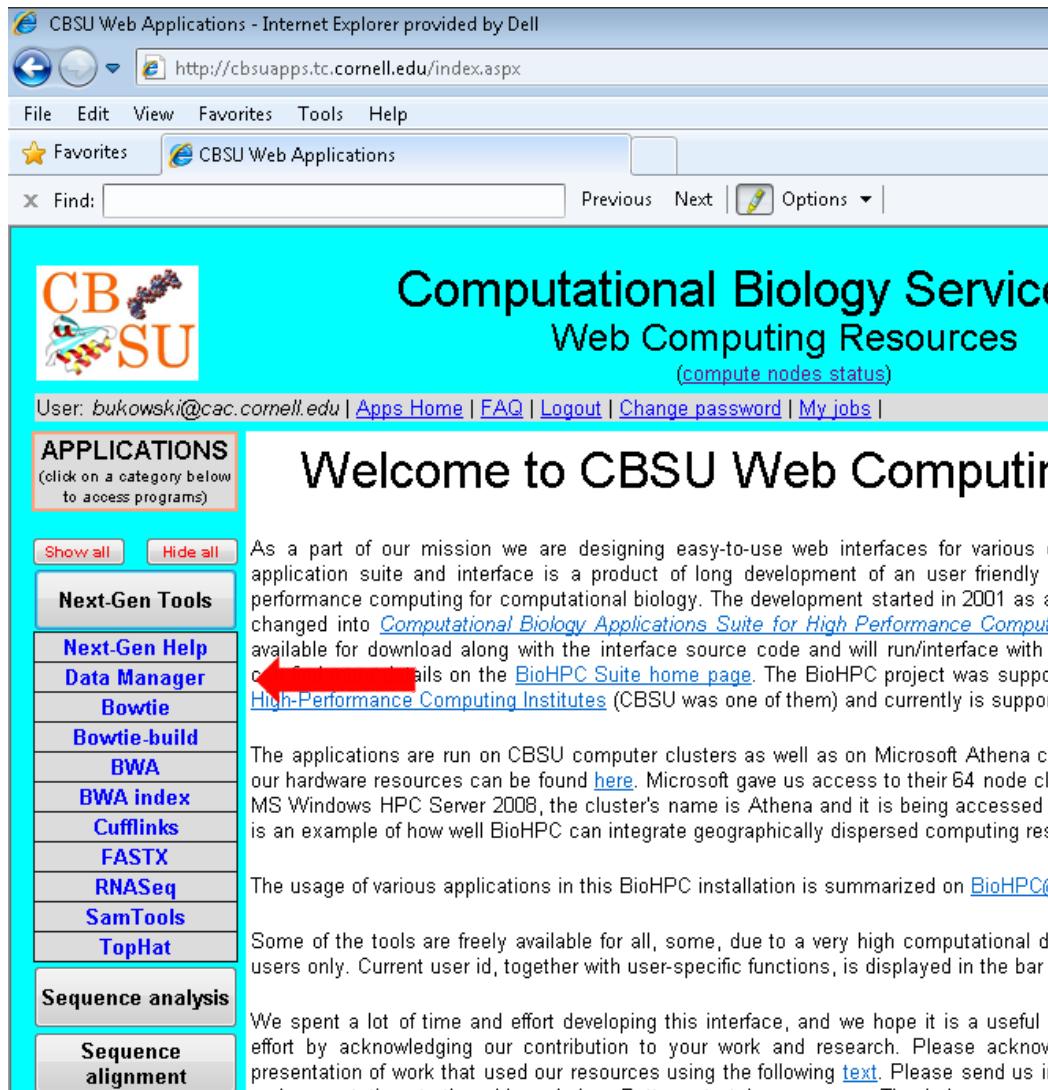
These files will show up in file selector dropdown lists in submission pages – no need to upload them before submitting a job!

Other files have to be first uploaded to BioHPC data store before you can use them as input to any BioHPC jobs. Examples of such files:

- “external” sequencing lanes, obtained outside of Cornell sequencing facility
- “private” reference genomes
- annotation files
- ...

**BioHPC Data Manager**: a web tool for listing, managing, upload, and download of files on BioHPC data store.

# Accessing BioHPC Data Manager interface



**Or go directly to**

<http://cbsuapps.tc.cornell.edu/Sequencing/seqmain.aspx>

# BioHPC Data Manager interface

g Home - Internet Explorer provided by Dell

http://cbsuapps.tc.cornell.edu/Sequencing/seqmain.aspx

View Favorites Tools Help

Sequencing Home

Next-Gen @ BioHPC

**Welcome to the BioHPC Next Generation sequencing data management and analysis pipelines interface**

This interface serves the purpose of storing and management of various data files to be used and processed by the Next Generation Sequencing analysis tools available through BioHPC. To run the tools on your data stored here, go to [the BioHPC interface page](#).

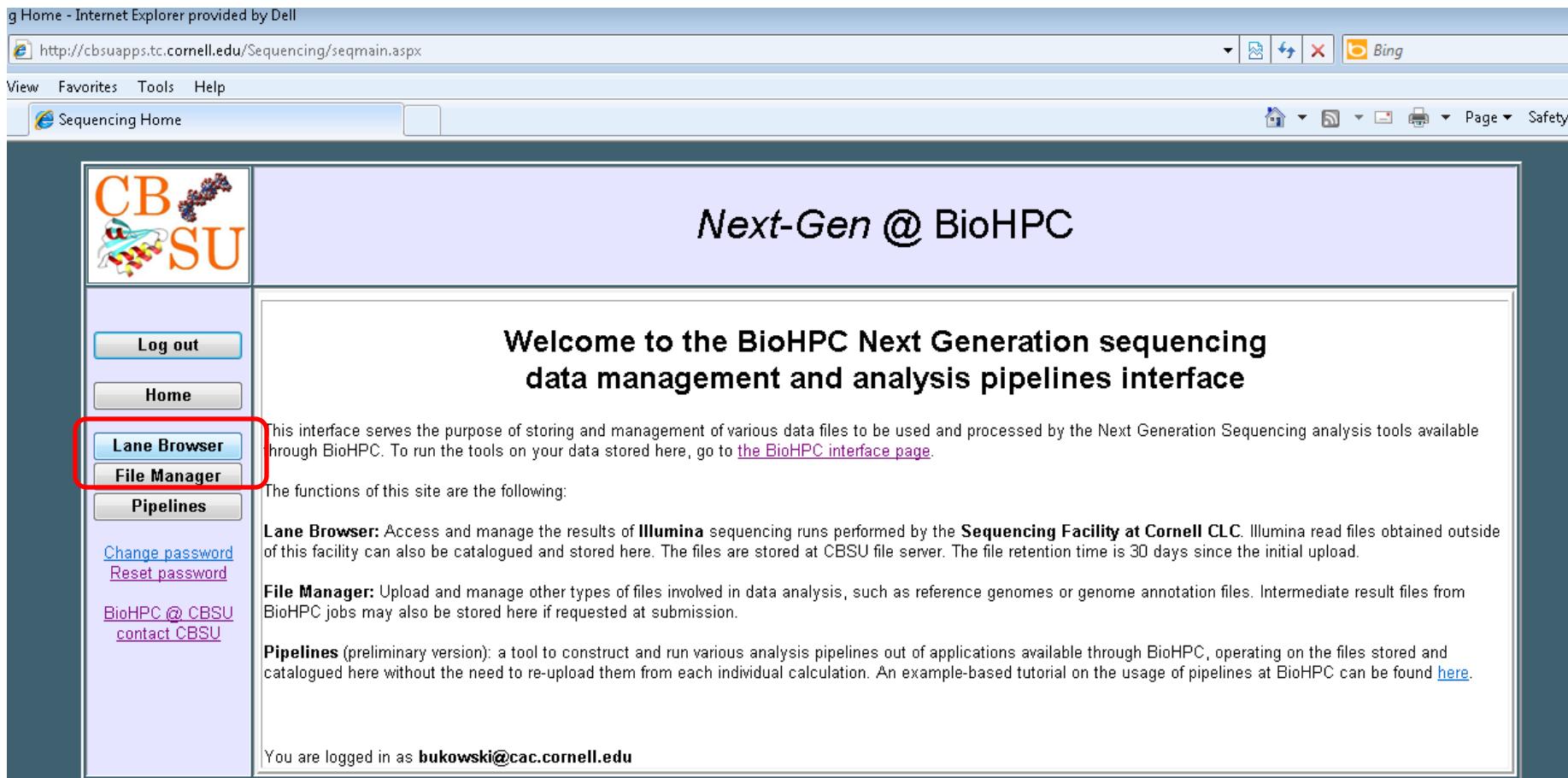
The functions of this site are the following:

**Lane Browser:** Access and manage the results of Illumina sequencing runs performed by the **Sequencing Facility at Cornell CLC**. Illumina read files obtained outside of this facility can also be catalogued and stored here. The files are stored at CBSU file server. The file retention time is 30 days since the initial upload.

**File Manager:** Upload and manage other types of files involved in data analysis, such as reference genomes or genome annotation files. Intermediate result files from BioHPC jobs may also be stored here if requested at submission.

**Pipelines** (preliminary version): a tool to construct and run various analysis pipelines out of applications available through BioHPC, operating on the files stored and catalogued here without the need to re-upload them from each individual calculation. An example-based tutorial on the usage of pipelines at BioHPC can be found [here](#).

You are logged in as **bukowski@cac.cornell.edu**



# BioHPC Data Manager: File Manager

Internet Explorer provided by Dell  
http://cbsuapps.tc.cornell.edu/Sequencing/refbrowser.aspx

Favorites Tools Help

File Manager

## FILE MANAGER

This page shows all registered files you have access to. The list can be sorted (click the column header of the field you want to sort over) and filtered (select the desired filters). Columns to be displayed can be selected by clicking on **Click to display/hide column selector**. After making column and/or filter selections, click **Refresh/Apply Filters** to activate the changes.

**Make Download Script** **Manage File Categories** **Process Selected Files** **Upload New File**

**Filter files by**

Category: All Format: All Status: All

File Name: \* File Description: \* Users: \*

**Click to display/hide column selector**

**Refresh/Apply Filters**

Select / Deselect all files on this page

**Click to share or change file category**

**Click to download**

**Look into source of the file**

File ID	File Name	Description	Category	Format	Compression	Source	Created	Users	Status	Size [bytes]
1659 <a href="#">(get file)</a>	my_test_aln_sorted_index	index of sorted test bam file	intermediate	bai	none	JOB 208039	3/5/2011 9:11:42 PM	bukowski@cac.cornell.edu (owner)	ready	2928
1658 <a href="#">(get file)</a>	my_test_alignment_sorted.bam	samtools output	intermediate	bam	none	JOB 208033	3/5/2011 9:02:42 PM	bukowski@cac.cornell.edu (owner)	ready	137131
1657 <a href="#">(get file)</a>	my_test_alignment.bam	This is a test alignment file produced at workshop	intermediate	bam	none	JOB 208030	3/5/2011 8:35:41 PM	bukowski@cac.cornell.edu (owner)	ready	228480
1656 <a href="#">(get file)</a>	indexes\Ecoli_NC_009600_BWAind.*	E coli BWA index	My BWA indexes	bwaind	none	JOB 208029	3/5/2011 8:26:46 PM	bukowski@cac.cornell.edu (owner)	ready	N/A

# BioHPC Data Manager: File Manager, cont.

fed by Dell  
nelli.edu/Sequencing/refbrowser.aspx

Help

Page Safety

1289	indexes\maize_agp_v2.*	Bowtie index for maize genome	Public Bowtie references	ebwt	none	uploaded	12/20/2010 7:38:38 PM	EVERYONE <a href="#">bukowski@tc.cornell.edu</a> (owner)	ready	N/A	
1267	indexes\maize_agp_v2.*	BWA index for maize genome	Public BWA references	bwaind	none	uploaded	12/15/2010 2:46:59 PM	<a href="#">bukowski@tc.cornell.edu</a> (owner) EVERONE	ready	N/A	
451 <a href="#">(get file)</a>	NC_009800.fna	Escherichia_coli_HS	reference genome	fasta	none	uploaded	7/2/2010 3:19:17 PM	<a href="#">bukowski@tc.cornell.edu</a> (owner) <a href="#">skirk@illumina.com</a> <a href="#">khaden@illumina.com</a> <a href="#">bukowski@cac.cornell.edu</a> <a href="#">dwilliamson@illumina.com</a> <a href="#">jp86@cornell.edu</a>	ready	4709943	
450 <a href="#">(get file)</a>	e_coli_10000snp.fq	Bowtie_Ecoli10000	Illumina lane	fastq	none	lane 483	7/2/2010 2:31:11 PM	<a href="#">bukowski@tc.cornell.edu</a> (owner) <a href="#">bukowski@cac.cornell.edu</a> <a href="#">dwilliamson@illumina.com</a>	ready	808890	
449 <a href="#">(get file)</a>	e_coli_1000_2.fq	Bowtie_Ecoli_paired-end	Illumina lane	fastq	none	lane 487	7/2/2010 2:29:44 PM	<a href="#">jp86@cornell.edu</a> <a href="#">bukowski@tc.cornell.edu</a> (owner) <a href="#">bukowski@cac.cornell.edu</a> <a href="#">khaden@illumina.com</a> <a href="#">dwilliamson@illumina.com</a> <a href="#">skirk@illumina.com</a>	ready	75890	
448 <a href="#">(get file)</a>	e_coli_1000_1.fq	Bowtie_Ecoli_paired-end	Illumina lane	fastq	none	lane 487	7/2/2010 2:29:44 PM	<a href="#">jp86@cornell.edu</a> <a href="#">bukowski@tc.cornell.edu</a> (owner) <a href="#">bukowski@cac.cornell.edu</a> <a href="#">khaden@illumina.com</a> <a href="#">dwilliamson@illumina.com</a> <a href="#">skirk@illumina.com</a>	ready	81890	
447 <a href="#">(get file)</a>	e_coli_1000.fq	Bowtie_EcoliSample	Illumina lane	fastq	none	lane 486	7/2/2010 2:26:51 PM	<a href="#">jp86@cornell.edu</a> <a href="#">bukowski@tc.cornell.edu</a> (owner) <a href="#">bukowski@cac.cornell.edu</a> <a href="#">khaden@illumina.com</a> <a href="#">dwilliamson@illumina.com</a> <a href="#">skirk@illumina.com</a>	ready	79890	
39 <a href="#">(get file)</a>	12345_314T3AAXX_s_3_2_sequence.txt.gz	Test Sample2	RB test categ 3	fastq	gz	lane 2	5/26/2010 5:41:19 PM	<a href="#">bukowski@cac.cornell.edu</a> (owner)	ready	1275586	

Cannot edit files I don't own

These files are public

These are Illumina lane files

# BioHPC Data Manager: managing file attributes

Log out

Home

Lane Browser

File Manager

Pipelines

[Change password](#)

[Reset password](#)

BioHPC @ CBSU  
[contact CBSU](#)

## EDIT FILE INFORMATION

You can change all editable fields, give access to an additional user (by selecting one from a dropdown list), deny access to a user currently with access (by checking the box next to this user's name), and change the file's owner (administrators only). After making all the changes, click **Submit Changes**. Multiple users can be added by repeating submission several times, each time with a new user selected from a dropdown list.

The **Category** field displays the category the file is currently assigned to. To change the file's category, select a new category from the dropdown list and click **Submit Changes**. If you need to create a new category, use the **Manage File Categories** button.

For files originating from a sequencing run (lanes) and files produced by BioHPC jobs, more information can be accessed by clicking on the lane or job ID.

To delete this file from BioHPC system, click the **Delete File** button. Sequencing lane files and files originating from a BioHPC job cannot be deleted using this method.

**Manage File Categories**

Expand to select change file category, if desired

Some fields are editable (depending on where the file came from)

Click after making changes

Expand to select additional user to share this file with other users

Parameter	Value
File ID:	1658
File Name:	my_test_alignment_sorted.bam
File Description:	samtools output
Category:	intermediate --- select new category ---
Format:	bam
Compression:	none
Source:	JOB
Lane/Job ID:	<a href="#">208033</a>
Version:	
Created:	3/5/2011 9:02:42 PM
Owner:	<a href="mailto:bukowski@cac.cornell.edu">bukowski@cac.cornell.edu</a>
Users:	<input checked="" type="checkbox"/> <a href="mailto:bukowski@cac.cornell.edu">bukowski@cac.cornell.edu</a> -- Add a user --
File status:	ready
File size [bytes]:	137131

**Submit Changes** **Delete this file** **Back to file list**

# Another way to share a file

Right-click and “Copy Shortcut”

1659 (get file)	my_test_aln_sorted_index	index of sorted bam file
1658 (get file)	my_test_alignment_sorted.bam	samtools outp
1657 (get file)	my_test_alignment.bam	This is a test alignment file produced at workshop

<http://cbsuapps.tc.cornell.edu/Sequencing/showseqfile.aspx?mode=http&cntrl=1052002694&refid=1658>

E-mail to a friend

# Download multiple files (to linux machine)

provided by Dell

tc.cornell.edu/Sequencing/refbrowser.aspx

Tools Help

Previous Next Options

CB SU

Log out

Home

Lane Browser

File Manager

Pipelines

change password

reset password

bioHPC @ CBSU

contact CBSU

"check" files of interest and click

Next-Gen @ BioHPC

FILE MANAGER

This page shows all registered files you have access to. The list can be sorted (click the column header of the field you want to sort over) and filtered (select the desired filters). Columns to be displayed can be selected by clicking on [Click to display/hide column selector](#). After making column and/or filter selections, click [Refresh/Apply Filters](#) to activate the changes.

Make Download Script Manage File Categories Process Selected Files Upload New File

Filter files by

Category: All Format: All Status: All

File Name: \* File Description: \* Users: \*

[Click to display/hide column selector](#)

[Refresh/Apply Filters](#)

Select / Deselect all files on this page

File ID	File Name	Description	Category	Format	Compression	Source	Version	Created	Users	Status	Size [bytes]
1659	my_test_aln_sorted_index	index of sorted test bam file	intermediate	bai	none	JOB 208039		3/5/2011 9:11:42 PM	bukowski@cac.cornell.edu (owner)	ready	2928
1658	my_test_alignment_sorted.bam	samtools output	intermediate	bam	none	JOB 208033		3/5/2011 9:02:42 PM	bukowski@cac.cornell.edu (owner)	ready	137131
1657	my_test_alignment.bam	This is a test alignment file produced at workshop	intermediate	bam	none	JOB 208030		3/5/2011 8:35:41 PM	bukowski@cac.cornell.edu (owner)	ready	228480
1656	indexes\Ecoli_NC_009800_BWAind.*	E coli BWA index	My BWA indexes	bwaind	none	JOB 208029		3/5/2011 8:26:46 PM	bukowski@cac.cornell.edu (owner)	ready	N/A

# Download multiple files, cont.

The screenshot shows a web browser window with the following details:

- Address Bar:** Dell | lu/Sequencing/refbrowser.aspx
- Toolbar:** Previous, Next, Options, Home, Refresh, Stop, Bind.
- Header:** Next-Gen @ BioHPC
- Left Sidebar (CBSU):** Log out, Home, Lane Browser, File Manager, Pipelines, Change password, Reset password, BioHPC @ CBSU, contact CBSU.
- Main Content Area:** A section titled "BATCH MODE DOWNLOAD" containing a Linux shell script fragment. The script uses the "wget" command to download multiple files from a specific URL pattern. A red oval highlights the command line:

```
    wget -q -O my_test_aln_sorted_index \
"http://cbsuapps.tc.cornell.edu/Sequencing/showseqfile.aspx?mode=http&cntrl=1904937385&refid=1659"
    wget -q -O my_test_alignment_sorted.bam \
"http://cbsuapps.tc.cornell.edu/Sequencing/showseqfile.aspx?mode=http&cntrl=1052002694&refid=1658"
    wget -q -O my_test_alignment.bam \
"http://cbsuapps.tc.cornell.edu/Sequencing/showseqfile.aspx?mode=http&cntrl=860458124&refid=1657"
```
- Bottom:** Back to file list button.

# BioHPC Data Manager: File categories

et Explorer provided by Dell  
[/cbsuapps.tc.cornell.edu/Sequencing/refbrowser.aspx?refid=1657](http://cbsuapps.tc.cornell.edu/Sequencing/refbrowser.aspx?refid=1657)

Favorites Tools Help  
File Manager

 Next-Gen @ Bio

**MANAGE FILE CATEGORIES**

Each file registered in File Manager belongs to one of the **file categories**. Grouping files into categories helps to organize files and filter the file listing. You can use a few categories pre-defined in the program, or create your own categories. This page helps you manage your categories.

To **rename** a category, edit text in column **CATEGORY**.  
To **delete** a category, check the box in column **DELETE**. The files from this category will be transferred to category "general". They **will not** be deleted.  
To **create** a new category, supply the name in the field **New Category**.  
After making all desired changes, click **Save Changes**.

You can only manage categories you created. The pre-defined categories (on top of the list) are non-editable.

Predefined categories

User-defined categories

Click after making changes

Check to remove category upon "Save Changes"

Specify name for new category, if desired

CATEGORY	DELETE
general	
Illumina lane	
intermediate	
genome annotation	
reference genome	
My BWA indexes	<input checked="" type="checkbox"/>
RB test categ 3	<input type="checkbox"/>
New Category:	

**Save Changes** **Back to file list**

# Uploading a new file

From file manager...

**Next-Gen @ BioHPC**

## FILE MANAGER

It can be sorted (click the column header of the field you want to sort over) and filtered (select the column selector. After making column and/or filter selections, click Refresh/Apply Filters to

**Manage File Categories** **Process Selected Files** **Upload New File**

Filter files by

Format: All Status: All

File Description: \* Users: \*

[Click to display/hide column selector](#)

**Refresh/Apply Filters**

... or directly from program's submission page

User: bukowski@cac.cornell.edu | Apps Home | FAQ | Logout | Change password | My jobs |

**APPLICATIONS**  
(click on a category below to access programs)

Show all Hide all

**Next-Gen Tools**

- Next-Gen Help**
- Data Manager**
- Bowtie
- Bowtie-build
- BWA
- BWA index
- Cufflinks
- FASTX
- RNASeq
- SamTools
- TopHat**

**Sequence analysis**

Sequence alignment

Population genetics

Protein structure

**TopHat @ BioHPC**  
Version v1.1.4  
Please send comments to [biohpc@cornell.edu](mailto:biohpc@cornell.edu).

TopHat is a fast splice junction mapper for RNA-Seq reads. It aligns RNA-Seq reads to the ultra high-throughput short read aligner Bowtie, and then analyzes the mapping results between exons. TopHat is a collaborative effort between the University of Maryland Center for Computational Biology and the University of California, Berkeley Departments of Mathematics and Biology.

To use TopHat, the Bowtie-formatted **index** of the reference genome database must be used using the **reference genome index file** selector (see below). If the list does not contain the required index file, you will have to first create this index from the FASTA file using the "Bowtie-build" tool, on the left, then "Bowtie-build" and then return to this page.

Calculations will be carried out on the BioHPC compute cluster at [CBSU](#). You will receive an email when it starts, and when it is finished. Output will be available via links on the right. For more information about this program and BioHPC interface in general, please see the [Questions](#) page.

Please acknowledge us in all publications and presentation of work that used our resources.

**Specify reference genome index file (required):**

-- select file(s) -- **Upload New File**

[filter file list](#) [more file details](#)

## File uploader (a Java applet)

Upload File - Internet Explorer provided by Dell  
<http://cbsuapps.tc.cornell.edu/Sequencing/regfilejava.aspx>

File Edit View Favorites Tools Help

★ Favorites Upload File

### Upload File

Your e-mail (as registered with BioHPC)  
bukowski@cac.cornell.edu

Messages:

Upload File  Upload Illumina Lane

Enter file to upload (with full path) [Browse](#)

Specify file format GTF

Specify file compression NONE

Specify file category genome annotation

Specify file description (optional)

**Upload File**

**Close Uploader**

Done

### To use file uploader:

- Need Java JRE 1.6 or newer + browser plug-in (standard)
- Accept the pop-up window
- Accept unverified digital signature (click "Run" when prompted)

# Uploading a new file

If uploading Illumina lane files, check the “Upload Illumina Lane” button

- Important for “parallel” files with paired-end reads

Upload File - Internet Explorer provided by Dell

http://cbsuapps.tc.cornell.edu/Sequencing/regfilejava.aspx

File Edit View Favorites Tools Help

Favorites Upload File

**Upload File**

Your e-mail (as registered with BioHPC)  
bukowski@cac.cornell.edu

Upload File     Upload Illumina Lane

Enter file to upload (with full path)

Specify file format

Specify file compression

Specify file category

Specify file description (optional)

Close Uploader

Done

Upload File - Internet Explorer provided by Dell

http://cbsuapps.tc.cornell.edu/Sequencing

File Edit View Favorites Tools Help

Favorites Upload File

**Upload Illumina Lane**

Your e-mail (as registered with BioHPC)  
bukowski@cac.cornell.edu

Upload File     Upload Illumina Lane

Enter file to upload (with full path)

Enter second file (if paired-end reads)

Specify file format

Specify file compression

Specify file description (optional)

Close Uploader

Done

Optional; if not provided, “unpaired” will be assumed

# Uploading a new file without Java applet

Next-Gen @ BioHPC

## FILE MANAGER

It can be sorted (click the column header of the field you want to sort over) and filtered (select the column selector. After making column and/or filter selections, click Refresh/Apply Filters to refresh the list.)

[Manage File Categories](#) [Process Selected Files](#) [Upload New File](#)

Filter files by

Format: All Status: All

File Description: \* Users: \*

[Click to display/hide column selector](#)

[Refresh/Apply Filters](#)

## UPLOAD AND REGISTER NEW FILE

er a new file in the File Manager.

obtained outside of the Cornell sequencing facility, use the Lane Browser rather than the Lane Browser for paired end sequencing run.

Two ways:

Java Applet. Requires Java run-time 1.6 or later and the corresponding browser plug-in to be installed. Configuration adjustments to run the applet. To use this route, click on the Use Java Applet button below.

Do NOT use Java (works only for small files, less than 50 MBytes in size), or register a file previously uploaded via ftp to our ftp server, then registered using this page (see text below for details).

[Use Java Applet](#) [Do NOT use Java](#)

[Back to file list](#)

For small files only (<50MB)

Larger files have to be first uploaded via ftp to our ftp server, then registered using this page (see text below for details)

## REGISTER NEW FILE

You can use this page to upload and register a new file in the File Manager.

Note: to register Illumina sequencing lanes obtained outside of the Cornell sequencing facility, use the Lane Browser rather than this page. This is especially important if the lane you want to register represents a paired end sequencing run.

Files up to 50 MB in size can be uploaded directly from your local machine (by selecting your local machine button and using the Local file selection browser).

Files larger than 50 MB should be first uploaded to the CBSU FTP server, to the directory /PBLAST (contact CBSU for upload credentials). Once the file has been uploaded to the CBSU FTP server, it can be registered using the form below by selecting the CBSU ftp server button and inserting the name of the file with path relative to PBLAST, if applicable, in the FTP file field (examples: If you deposited your file called my\_sequence.fasta directly in the /PBLAST directory, specify the name as my\_sequence.fasta; if, instead, you created a subdirectory my\_sub\_dir in PBLAST and deposited your file there, enter my\_sub\_dir/my\_sequence.fasta as the file name [note the use of "V" rather than "/"]).

Regardless of the source of the file, please specify the file category (a new category may be created using the Manage File Categories button), format, and compression type. You can also insert a short description of the file.

After specifying all fields, click Submit to register the file in File Manager. The files will be transferred asynchronously, which may take some time - typically about 15 minutes. You will obtain a confirmation e-mail once the files are transferred and registered. You may then want to log back in to the CBSU FTP server and remove your files (although this step is not strictly necessary).

[Manage File Categories](#)

Parameter	Value
Source:	<input checked="" type="radio"/> your local machine <input type="radio"/> CBSU ftp server
Category:	<input type="button" value="... select category ..."/>
Format:	<input type="button" value="... select format ..."/>
Compression:	<input type="button" value="... select compression ..."/>
Description:	<input type="text"/>
Local file:	<input type="text"/> <input type="button" value="Browse..."/>
FTP file:	<input type="text"/>

[Submit](#) [Back to file list](#)

# Lane Browser: tool to manage Illumina lane files

**Lane Browser** is complementary to File Manager (after all, Illumina read files are just files and as such they are visible in File Manager)

**Lane Browser** displays some lane-specific information, not available through File Manager

The screenshot shows a web browser window titled "BROWSE MY LANES". The left sidebar has links for Log out, Home, Lane Browser (which is selected and highlighted with a red box), File Manager, Pipelines, Change password, and Reset password. Below these are links for BioHPC @ CBSU and contact CBSU. A red box highlights the "Lane Browser" link in the sidebar.

A red box also highlights the "Lane Browser" tab in the main content area. The main content area has a heading "BROWSE MY LANES" and a descriptive text about lane management. It includes three buttons: "Make download script", "Process selected lanes", and "Register external lane". Below these are filters for Status (Active), Sample name, and Users, followed by an "Apply Filters" button. A red box highlights the "Lane ID" column header in the table below.

The table lists three lanes:

Lane ID	Run ID	Run Name	Lane#	Type	Sample Name	Status	Annotations			Users	Expires	Order#
10	4	090203_HWI-EAS339_30RF7AAXX	4	Standard	actual project from datarig	ready	Parameter	This Lane	Ctrl Lane	bukowski@cac.cornell.edu jv2@cornell.edu (owner) jp86@cornell.edu pschweitzer1@twcny.rr.com bukowski@udel.edu	8/25/2012	12345
2		090504_HWI-EAS339_314T3AAXX	3	PhiX	Test Sample2	ready	Parameter	This Lane	Ctrl Lane	bukowski@cac.cornell.edu (owner) maryh@cac.cornell.edu	8/20/2012	12345
488		N/A	N/A	N/A	E coli sequencing	ready	--- NO ANNOTATIONS ---			bukowski@cac.cornell.edu dwilliamson@illumina.com bukowski@tc.cornell.edu (owner)	1/1/2020	N/A

Red boxes highlight specific elements: "Click on ID to manage access" points to the Lane ID column; "Click on '(files)' to download" points to the "(files)" link in the first row; "Lanes from Cornell facility (uploaded automatically)" points to the first two rows; and "‘external’ lanes (uploaded by user)" points to the third row.

# BioHPC Data Manager: sharing a lane

Parameter	Value												
Run ID:	2												
Run Name:	090504_HWI-EAS339_314T3AAXX												
Lane#:	3												
Analysis Software:	Bustard 1.3.2												
Type:	PhiX												
Sample Name:	Test Sample2												
Status:	ready												
Annotations:	<table><thead><tr><th>Parameter</th><th>This Lane</th><th>Ctrl Lane</th></tr></thead><tbody><tr><td>Length</td><td>unknown</td><td>unknown</td></tr><tr><td>Clusters_raw</td><td>1.0M,1.0M</td><td>13.5M,13.5M</td></tr><tr><td>Clusters_PF</td><td>23.1K,23.1K</td><td>9.8M,9.8M</td></tr></tbody></table>	Parameter	This Lane	Ctrl Lane	Length	unknown	unknown	Clusters_raw	1.0M,1.0M	13.5M,13.5M	Clusters_PF	23.1K,23.1K	9.8M,9.8M
Parameter	This Lane	Ctrl Lane											
Length	unknown	unknown											
Clusters_raw	1.0M,1.0M	13.5M,13.5M											
Clusters_PF	23.1K,23.1K	9.8M,9.8M											
Users:	<p><input type="checkbox"/> <a href="#">bukowski@cac.cornell.edu</a> (owner) <input type="button" value="send link"/></p> <p><input checked="" type="checkbox"/> <a href="#">maryh@cac.cornell.edu</a> <input type="button" value="send link"/></p> <p>-- Add a user -- <input type="button" value="▼"/></p>												
Lab:	N/A												
Order#:	12345												
Expires:	8/20/2012 12:00:00 AM												

[Download files for this lane](#)

Click to access file download page

Check to remove user on "Submit Changes"

Expand to select user from list

Click after making access changes

Click to send link to user

# Another way to share a lane a lane

The screenshot shows a Microsoft Internet Explorer window with a red oval highlighting the URL bar. The URL is <http://cbsuapps.tc.cornell.edu/Sequencing/showSeqfile.aspx?mode=outlist&cntrl=1234567&laneid=2>. The page title is "showSeqfile - Internet Explorer provided by Dell". The main content displays sequencing results for sample "Test Sample2".

**Sequencing results for sample "Test Sample2"**

Parameter	Value												
Run Name:	090504_HWI-EAS339_314T3AAXX												
Lane#:	3												
Analysis Software:	Bustard 1.3.2												
Sample Name:	Test Sample2												
Lane Annotations:	<table><thead><tr><th>Parameter</th><th>This Lane</th><th>Ctrl Lane</th></tr></thead><tbody><tr><td>Length</td><td>unknown</td><td>unknown</td></tr><tr><td>Clusters_raw</td><td>1.0M,1.0M</td><td>13.5M,13.5M</td></tr><tr><td>Clusters_PF</td><td>23.1K,23.1K</td><td>9.8M,9.8M</td></tr></tbody></table>	Parameter	This Lane	Ctrl Lane	Length	unknown	unknown	Clusters_raw	1.0M,1.0M	13.5M,13.5M	Clusters_PF	23.1K,23.1K	9.8M,9.8M
Parameter	This Lane	Ctrl Lane											
Length	unknown	unknown											
Clusters_raw	1.0M,1.0M	13.5M,13.5M											
Clusters_PF	23.1K,23.1K	9.8M,9.8M											
Order#:	12345												
Expiration Date:	8/20/2012												

Files will be available for download until **8/20/2012 (532 days left)**

File (click to download)	Size [bytes]	MD5 sum
<a href="#">12345_314T3AAXX_s_3_1_sequence.txt.gz</a>	1,227,597	bb86e742186af0922378a99a446a81b4
<a href="#">12345_314T3AAXX_s_3_2_sequence.txt.gz</a>	1,275,586	e7471d6373956297890a6c6a4f41bb99

Copy the URL and e-mail it to the person you want to share the lane files with

# Directions for BioHPC Web Resources

## Simplicity vs. flexibility trade-off

- Simplicity: implement a few standardized, “packaged” pipelines (e.g., CBSU RNA-Seq, BWA), where complex, multi-step and multi-tool procedures are launched at a click of a button
  - Limited user customization possibilities
  - Standard procedures not always available in active research environment
- Flexibility: implement a lot of “one-step” tools (samtools, FASTX) and let the user connect them into pipelines (**Pipeline Manager**, see [http://cbsuapps.tc.cornell.edu/doc/Pipelines\\_Manual.pdf](http://cbsuapps.tc.cornell.edu/doc/Pipelines_Manual.pdf))
  - Large number of web interfaces need to be maintained for multiple tools
  - Learning curve involved in web-based pipeline construction becomes steeper
  - “Cut the middleman” and learn Linux instead?

**Suggestions welcome**