Software

Installed Application Software:

Here is a list of all 64-bit applications currently installed on the research cluster. It is policy to retain only the two most recent versions.

Ansys

Ansys is a suite of finite element based applications that provide real-world simulations of structural, thermal, electromagnetic and fluid-flow behavior of 3-D product. All Ansys products integrate with CAD environments. Ansys's **Polyflow** is now available. This is normally access via the **WorkBench** interface.

Fluent is a Computational Fluid Dynamics (CFD) software package commonly used in engineering education for research in fluid mechanics. ANSYS purchased Fluent recently and has incorporated Fluent functionality into Ansys on an ongoing basis. The Ansys WorkBench product is how you access this functionality. Note Fluent 2d and 3d and Icepak products are available outside of WorkBench.

Abaqus
Abaqus is a suite of applications used by many in the engineering community for the analysis of multi-body dynamics problems that aid the medical, automotive, aerospace, defense, and manufacturing community.

Comsol

Comsol is specifically designed to easily couple transport phenomena, including computational fluid dynamics (CFD) as well as mass and energy transport to chemical-reaction kinetics and process-related modeling. Licensed Modules include: MultiPhysics, Chemical Engineering, Acoustics, Structural Mechanics, Script.

Chimera

Chimera is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. High-quality images and animations can be generated.

Cubit
CUBIT is a Geometry and Mesh Generation Toolkit

Desmond
Desmond is a software package developed at D. E. Shaw Research to perform high-speed molecular dynamics simulations of biological systems.

FFmpeg
FFmpeg is a complete, cross-platform solution to record, convert and stream audio and video. It includes libavcodec - the leading audio/video codec library.

FE-Safe
The Abaqus add-on product FE-Safe is a highly effective tool for fatigue analysis of Finite Element models.

gnuplot
Gnuplot supports many types of plots in either 2D and 3D. It can draw using lines, points, boxes, contours, vector fields, surfaces, and various associated text. It also supports various specialized plot types.

ggobi
GGobi is an open source visualization program for exploring high-dimensional data. It provides highly dynamic and interactive graphics such as tours, as well as familiar graphics such as the scatterplot, barchart and parallel coordinates plots. Plots are interactive and linked with brushing and identification.

Gromacs
GROMACS is a versatile package to perform molecular dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles.
It is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of complicated bonded interactions, but since GROMACS is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, e.g. polymers.

GAMESS
GAMESS is a program for ab initio molecular quantum chemistry.

GMT
GMT is an open source collection of ~60 tools for manipulating geographic and Cartesian data sets.

Imagemagick

ImageMagick® is a software suite to create, edit, and compose bitmap images. It can read, convert and write images in a variety of formats (over 100) including DPX, EXR, GIF, JPEG, JPEG-2000, PDF, PhotoCD, PNG, Postscript, SVG, and TIFF. Use ImageMagick to translate, flip, mirror, rotate, scale, shear and transform images, adjust image colors, apply various special effects, or draw text, lines, polygons, ellipses and Bézier

curves.

LSF

Platform Computing, Inc.'s **LSF** (*Load Sharing Facility*) software is a distributed load sharing and batch queuing suite of applications that can dispatch user requests to compute nodes in accordance with a Tufts-defined policy. It manages, monitors, and analyzes resources and load on the cluster. Platform LSF is layered in a way that allows it to sits on top of and extend the operating system services, speaking to the competing needs of resource management on the cluster. LSF commands must be used to submit batch jobs and assign interactive jobs to processors. **bsub** and **lsrun** are the usual command tools for this. It's important to note that cluster compute nodes are the only targets under LSF control. Jobs are not submitted to computers outside of the cluster. For more information about LSF command usage and job submission, you can read the **man** pages (*example: type **man lsrun** at the cluster prompt*). Most commands begin with the letter 'b' such as: **bsub**, **bkill**, **lsrun**, and **bjobs**.

Matlab

MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis, and numerical computation. Using MATLAB, you can solve technical computing problems faster than with traditional programming languages, such as C, C++, and Fortran. Extensive documentation and tutorials are provided within Matlab. The following Matlab toolboxes are licensed, however for an up to date list, type **ver** within Matlab's command line:

MATLAB, Simulink,Control System,Curve Fitting,Distributed Computing,Financial,Fuzzy Logic,Image Processing,MATLAB Compiler,Neural Network,Optimization,Partial Differential Equation,Real-Time Workshop,Signal Processing,Simulink Control Design,Simulink 3D Animation,Spline,Statistics,Symbolic Math,System Identification,Virtual Reality,Wavelet Toolbox,Bioinformatics,Simbiology

MATLAB Compiler Runtime (MCR)

The MATLAB Compiler Runtime (MCR) is a standalone set of shared libraries that enables the execution of compiled MATLAB applications or components without using MATLAB licensing. This is useful for running a large number of jobs that might exceed the license counts.

Maple

Maple is a well known environment for mathematical problem-solving, exploration, data visualization, and technical authoring. In may ways it is similar to Mathematica and Matlab.

MCEE

MCCE (Multi-Conformation Continuum Electrostatics) is a biophysics simulation program combining continuum electrostatics and molecular mechanics.

Mathematica

Mathematica, advertised as a one stop for technical work that integrates a numeric and symbolic computational engine, graphics system, programming language, documentation and advanced connectivity to other applications. Not only does this application have parallel functionality built into it from the ground up. The Wolfram Mathematica web site has extensive documentation, including numerous detailed tutorials.

NGSPICE

Spice is a general-purpose electric circuit simulation program for nonlinear dc, nonlinear transient and linear ac analysis.

NCAR

(The NCAR Command Language (NCL), a product of the Computational & Information Systems Laboratory at the National Center for Atmospheric Research (NCAR) and sponsored by the National Science Foundation, is a free interpreted language designed specifically for scientific data processing and visualization.

Namd

NAMD is a parallel molecular dynamics code designed for high-performance simulation of large biomolecular systems.

PetSc

PetSc is a suite of data structures and routines for the scalable (parallel) solution of scientific applications modeled by partial differential equations. It employs the MPI standard for parallelism. Note, another useful **link** to visit. PetSc has a complex build environment that allows one to include various applications. We encourage interested users to build the suite of interest in their home directory. Installation instructions for a particular set of applications can be found **here**

Paraview

ParaView is a multi-platform visualization application designed to visualize large data sets.

RATS

RATS(Regression Analysis of Time Series) is a leading econometrics and time-series analysis software package.

R

R is a widely available object oriented statistical package. The current list of installed packages, to numerous to list, can be found in directory /opt/shared/R/2.15.0-rhel6/lib64/R/library/.

This represents an augmented base installation suitable for most routine tasks, however not all available packages as found on the R web site are installed. If some other R package is needed, please make a software installation request as outlined above. However, for many people use of **RS tudio** will allow for local installations in your home directory. Extensive user documentation and tutorials are available on the R web site. There are many texts as well, here is nice **example**

Note **BioConductor** and many genetics related R packages are installed as well.

RStudio
RStudio is a free and open source integrated development environment(IDE) for R.

Stata

Stata SE is an integrated statistical package for Windows, Macintosh, and Unix platforms. More than just a statistical package, Stata is also a full data-management system with complete statistical and graphical capabilities. It features both X-window and text user interfaces.

SAS
Sas is general purpose statistics package. Installed is the Education Analytical Suite of programs. These include:

• Base SAS
• SAS Bridge for ESRI
• SAS Enterprise Guide
• SAS Integration Technologies
• SAS/ACCESS
• SAS/AF
• SAS/ASSIST
• SAS/CONNECT
• SAS/EIS
• SAS/ETS
• SAS/FSP
• SAS/GRAPH
• SAS/IML
• SAS/INSIGHT
• SAS/LAB
• SAS/OR
• SAS/QC
• SAS/SECURE 168-bit
• SAS/SHARE
• SAS/STAT
SAS Enterprise Guide
SAS/Genetics

TecPlot
Tecplot 360 is a numerical simulation and CFD visualization software that combines vital engineering plotting with advanced data visualization into one tool. It allows you to quickly plot and animate all your data exactly the way you want, as well as analyze complex data, arrange multiple layouts, and communicate your results with professional images and animations.

Weka

Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

WPP

WPP is a parallel computer program for simulating time-dependent elastic and viscoelastic wave propagation, with some provisions for acoustic wave propagation. *WPP* solves the governing equations in displacement formulation using a node-based finite difference approach on a Cartesian grid. *WPP* implements substantial capabilities for 3-D seismic modeling,

Public domain Matlab Toolboxes

Edge

Wavelab

Installed HPC Math Libraries

See **Standalone Math Libraries** link.

Bioinformatic cluster installed software tools

User application specific documentation is found on the following sites. Tufts cluster Application FAQ section details how to start the application.

SimBiology and **BioInformatic** Matlab Toolboxes.
These are licensed under Tufts Network concurrent license. Both toolboxes are included in the matlab user session. For additional examples and demos see:

Working with **Illumina data**

Working with **454 data**

All Bioinformatics demos can be accessed on this **demo page**

MATLAB Central **File Exchange** for user contributed matlab scripts.

blast
BLAST+ is a new suite of BLAST tools that utilizes the NCBI C++ Toolkit. The BLAST+ applications have a number of performance and feature improvements over the legacy BLAST applications.

fastx
FASTX-Toolkit is a collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing.

intrapid
Software for Genome-wide association studies (**GWAS**).

mrbayes
MrBayes is a program for the Bayesian estimation of phylogeny.

MAQ
Genetic Mapping and Assembly tools.

ancestrymap
ANCESTRYMAP screens through the genome in a recently mixed population such as African Americans, searching for segments with increased ancestry from one of the ancestral populations, which can indicate the position of disease genes

eigensoft
The EIGENSOFT package combines functionality from our population genetics methods (Patterson et al. 2006) and our EIGENSTRAT stratification method (Price et al. 2006). The EIGENSTRAT method uses principal components analysis to explicitly model ancestry differences between cases and controls along continuous axes of variation; the resulting correction is specific to a candidate marker's variation in frequency across ancestral populations, minimizing spurious associations while maximizing power to detect true associations.

FigTree
FigTree is designed as a graphical viewer of phylogenetic trees.

Structure
The program structure is a software package for using multi-locus genotype data to investigate population structure.

fbat
FBAT implements a broad class of Family Based Association Tests, adjusted for population admixture.

metal
The METAL software is designed to facilitate meta-analysis of large datasets (such as several whole genome scans) in a convenient, rapid and memory efficient manner.

haploview
Haploview computes single locus and multi-marker haplotype association tests. Haploview provides a framework for permuting your association results in order to obtain a measure of significance corrected for multiple testing bias.

impute
IMPUTE is a program for estimating ("imputing") unobserved genotypes in SNP association studies.

merlin
MERLIN uses sparse trees to represent gene flow in pedigrees and is one of the fastest pedigree analysis packages around.

plink
PLINK is a free, open-source whole genome association analysis toolset,
designed to perform a range of basic, large-scale analyses in a computationally efficient manner. The focus of PLINK is purely on analysis of genotype/phenotype data, so there is no support for steps prior to this (e.g. study design and planning, generating genotype or CNV calls from raw data). Through integration with gPLINK and Haploview, there is some support for the subsequent visualization, annotation and storage of results.

pbat
Tools for family-based association tests (FBAT)

pedcheck
A program for detecting marker typing incompatibilities in pedigree data.

solar
SOLAR is an extensive, flexible software package for genetic variance components analysis, including linkage analysis, quantitative genetic analysis, SNP association analysis (QTN and QTLD), and covariate screening. Operations are included for calculation of marker-specific or multipoint identity-by-descent (IBD) matrices in pedigrees of arbitrary size and complexity, and for linkage analysis of multiple quantitative traits and/or discrete traits which may involve multiple loci (oligogenic analysis), dominance effects, household effects, and interactions.

MACH

MACH is a Markov Chain based haplotyper. It can resolve long haplotypes or infer missing genotypes in samples of unrelated individuals.

Beagle
BEAGLE is a state of the art software package for analysis of large-scale genetic data sets with hundreds of thousands of markers genotyped on thousands of samples. This software is not installed system-wide, as it does not require such. Instead download it by using the following command: wget http://www.stat.auckland.ac.nz/~bbrowning/beagle/beagle.jar

IMa2
This program implements a method for generating posterior probabilities for complex demographic population genetic models for two or more populations.

Velvet
This is a Sequence assembler program for very short genetic reads. Both versions are installed.

R
In addition, many genetics R based software packages are installed. See above **R** listing.

Cufflinks
Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples.

Bowtie
Bowtie is an ultrafast, memory-efficient short read aligner.

Tophat
TopHat is a fast splice junction mapper for RNA-Seq reads.

BioPerl
Perl code which is useful in biology. Bioperl provides software modules for many of the typical tasks of bioinformatics programming. These include:

- Accessing sequence data from local and remote databases
- Transforming formats of database/ file records
- Manipulating individual sequences
- Searching for similar sequences
- Creating and manipulating sequence alignments
- Searching for genes and other structures on genomic DNA
- Developing machine readable sequence annotations

phred, phrap, consed, etc...
The phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base. See the site for descriptions of the other bundled software.

GATK
GATK was designed to simplify the process of developing efficient, robust tools for working with NGS data, and currently supports in a single integrated framework Solexa, SOLiD, 454, Complete Genomics, and Sanger sequencer data.

IGV
The Integrative Genomics Viewer (IGV) is a high-performance visualization tool for interactive exploration of large, integrated datasets.

polyphred
PolyPhred is a program that compares fluorescence-based sequences.

mothur
Mothur addresses the bioinformatics needs of the microbial ecology community. Functionality such as dotur, sons, treeclimber, s-libshuff, unifrac, and other algorithms are provided. In addition to improving the flexibility of these algorithms, a number of other features including calculators and visualization tools offers the ability to go from raw sequences to the generation of visualization tools.

diyabc
Approximate Bayesian Computation for inference on population history using molecular markers.

QIIME
QIIME is an open source software package for comparison and analysis of microbial communities.

RetroSeq
RetroSeq is a bioinformatics tool that searches for mobile element insertions from aligned reads in a BAM file and a library of reference transposable elements.

BedTools

Exonerate

SAMtools

Tufts Medical School Bioinformatic cluster installed software tools

A local software repository is available to meet the dynamic needs of individual Medical researchers. This effort is maintained by Lax Iyer(lax.iyer@tufts.edu), Joshua Ainsley(Joshua.Ainsley@tufts.edu), and Gavin Schnitzler(GSchnitzler@tuftsmedicalcenter.org). Please contact them to address specific software needs or requests.

Packages for Next Generation Sequencing analysis have been set up in filesystem **/cluster/tufts/ngsp/ngsp/bin/**

Programs in this directory can be accessed using the module command,

> **module add ngsp**

which adds the path **/cluster/tufts/ngsp/ngsp/bin** to a user's PATH.

Available software:
This collection is always growing. To view what is actually there, do the following:
> ls /cluster/tufts/ngsp/ngsp/bin | more
Use the space bar to page the listing.

Bowtie indices and annotations from Illumina for the Mouse

Illumina annotation and index information can be found in:

/cluster/tufts/ngsp/ngsp/Mus_musculus/Ensembl
/cluster/tufts/ngsp/ngsp/Mus_musculus/UCSC

For ENSEMBL:
The annotations is at:

/cluster/tufts/ngsp/ngsp/Mus_musculus/Ensembl/NCBIM37/Annotation/Archives/archive-2013-03-06-18-55-12/Genes/genes.gtf

and the bowtie indices are in:
/cluster/tufts/ngsp/ngsp/Mus_musculus/Ensembl/NCBIM37/Sequence/BowtieIndex

For UCSC: The corresponding files are in directories named Sequence and Annotation

Remember: UCSC used "chr" but ENSEMBL uses on number to name chromosomes.

Bowtie is an ultrafast, memory-efficient short read aligner. It aligns short DNA sequences (reads) to the human genome at a rate of over 25 million 35-bp reads per hour. Bowtie indexes the genome with a Burrows-Wheeler index to keep its memory footprint small: typically about 2.2 GB for the human genome (2.9 GB for paired-end) **Bowtie documentation**

TopHat is a program that aligns RNA-Seq reads to a genome in order to identify exon-exon splice junctions. It is built on the ultrafast short read mapping program Bowtie. TopHat runs on Linux and OS X. **TopHat documentation**

Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples. It accepts aligned RNA-Seq reads and assembles the alignments into a parsimonious set of transcripts. Cufflinks then estimates the relative abundances of these transcripts based on how many reads support each one, taking into account biases in library preparation protocols. **Cufflinks documentation**

SAMtools: SAM Tools provide various utilities for manipulating alignments in the SAM format, including sorting, merging, indexing and generating alignments in a per-position format **SAMtools documentation**

BamTools provides both a programmer's API and end-user's toolkit for handling BAM files **BamTools documentation**

Mothur:This project seeks to develop a single piece of open-source, expandable software to fill the bioinformatics needs of the microbial ecology community **Mothur documentation**

RSEM is a software package for estimating gene and isoform expression levels from RNA-Seq data. The new RSEM package (rsem-1.x) provides an user-friendly interface, supports threads for parallel computation of the EM algorithm, single-end and paired-end read data, quality scores, variable-length reads and RSPD estimation. It can also generate genomic-coordinate BAM files and UCSC wiggle files for visualization. In addition, it provides posterior mean and 95% credibility interval estimates for expression levels **RESM documentation**

CEAS
Software for Cis-regulatory Element Annotation System.
Info at: **CEAS**

Installed in /cluster/tufts/ngsp/ngsp/bin
See the **lib** and **bin** directories in this directory.

You have to modify the PYTHONPATH environment variable:

set PYTHONPATH = /cluster/tufts/ngsp/ngsp/bin/lib/python2.6/site-packages/

Some supporting files including test data are available at
/cluster/tufts/ngsp/ngsp/CEAS

Flash
Fast Length Adjustment of Short reads is a very fast and accurate software tool to merge paired-end reads from next-generation sequencing

experiments. Website for **Flash**.

PeakSplitter
Software for Subdivision of ChIP-seq/ChIP-chip regions into discrete signal peaks. Website for PeakSplitter*.

Installation in cluster directory:
/cluster/tufts/ngsp/ngsp/PeakSplitter_Cpp/PeakSplitter_Linux64