

# Software Resources

Many popular software packages are installed and available as modules. Frequently, different versions are available and you can use an older version. The default in most cases is the latest version. For example, "module load julia" will provide Julia version 1.6.1 (at the time of writing), whereas "module load julia/1.4.1" will make an older version available.

You may request that applications be installed if they have general usefulness - <https://it.arizona.edu/hpc-software-install-request>.

**Puma, Ocelote, and ElGato** are built on CentOS 7 which is the foundation for all compilers, libraries, and applications. The same software applications are available on all three supercomputers as they are presented from a unified filesystem.

Most software applications are installed as modules and are loaded using module commands. The short explanation is that loading a module sets the path so that you do not need to know where it is installed. For example, after doing a "module load namd", you can invoke "namd3". The module load will set other required elements like loading Cuda. More information is at [this location](#).

Most software that needed to be compiled has used GCC version 8.3. It was the latest available when the software builds were started in 2020. Some other packages are built with the Intel compiler. Both are available to compile your own code. Note that GCC is loaded by default along with OpenMPI 3 (named gnu8 and openmpi3 respectively).

## Installed Software



- The list here is usually incomplete because of the frequent changes.
- Some modules are in the operating system also, but are too downlevel, like curl or libtool.
- The current list of modules is available on each login node with the command "module avail".
- Module commands are [available here](#).

Software	Description
Abaqus	Finite element analysis
Abyss	Parallel assembler for short read sequence data
Amira	Powerful, multifaceted 2D–5D software for visualization, processing and analysis of microscopy imaging serving Life and Biomedical Sciences.
Anaconda	Platform for data science and machine learning
Ansys	Licensed: General purpose finite element modeling package
Ant	JAVA build tool
AOCL	AMD Optimized CPU Libraries. Includes BLIS, libFLAME, FFTW, LibM, ScaLAPACK
Atlas	Automatically Tuned Linear Algebra Software (ATLAS). C and Fortran77 interfaces to a portably efficient BLAS implementation, as well as a few routines from LAPACK
Augustus	Gene prediction and is required for Maker
Autoconf	GNU tool for producing configure scripts for building, installing and packaging software on computer systems where a Bourne shell is available
Bamtools	C++ API and toolkit for analyzing and managing BAM files
Bbmap	Short read aligner for DNA and RNA-seq data
Bcftools	A set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF
Beagle-lib	Phylogenetics - works with Beast
Beast	Bayesian analysis of molecular sequences
Bedtools	Utilities for comparing, summarizing and intersecting genomic features in the UCSC Genome Browser BED format
Biocontainers	<a href="#">Biocontainers</a> is a registry of Biology tools that can be pulled from a Docker container into a <a href="#">Singularity container</a> .
Bison	General purpose parser generator that converts a grammar description for an LALR(1) context-free grammar into a C program to parse that grammar.
Blas	Basic Linear Algebra Subprograms

Blast	Search tool that finds regions of local similarity between nucleotide or protein sequences
Blat	Alignment tool like BLAST, but structured differently
Blender	3D Visualization software
Boost	Peer reviewed C++ source libraries
Bowtie2	Ultrafast, memory-efficient tool for aligning sequencing reads to long reference sequences.
BWA	Fast light-weight tool that aligns relatively short sequences to a sequence database
Bzip2	Compresses data in blocks of size between 100 and 900 kB and uses the <a href="#">Burrows–Wheeler transform</a> to convert frequently-recurring character sequences into strings of identical letters
Caffe	Deep learning framework made with expression, speed, and modularity
Canu	Fork of the Celera Assembler designed for high-noise single-molecule sequencing
CCP	Cisco Configuration Professional is a GUI based device management tool for Cisco access routers.
CDO	Climate Data Operator is a collection of command-line Operators to manipulate and analyze Climate and NWP model Data
Cell Ranger	A set of analysis pipelines that process Chromium single-cell data to align reads, generate feature-barcode matrices, perform clustering.
Cern-Root	A modular scientific software framework
Cfitsio	A library of subroutines for data files
Chapel	Programming language designed for productive parallel computing at scale
Cluster	The open source clustering software implements the most commonly used clustering methods for gene expression data analysis
ClusterShell	ClusterShell is an event-driven open source Python library, designed to run local or distant commands in parallel on server farms or on large Linux clusters
Cmake	Tool to control the compilation process. "cmake" comes with the operating system, but it is an old version 2. Likely you should "module load cmake" which will get you a much newer release of version 3.
Comsol	Modeling and simulating physics-based problems (licensed)
Contrib	<a href="#">Adds user supported software</a> to your module path. On Ocelote this is /unsupported
Coot	Macromolecular model building
CP2K	Open Source Molecular Dynamics
Cplex	IBM optimization models, combining leading solver engines with a tightly integrated IDE and modeling language
Cuda	Parallel computing platform and API model for Nvidia GPU's
Cufflinks	Assembles transcripts, estimates their abundances, tests for differential expression and regulation in RNA-Seq samples
Curl	Computer software project providing a library and command-line tool for transferring data using various protocols
dbg2olc	Efficient assembly of large genomes
dealii	deal.II is a C++ program library targeted at the computational solution of partial differential equations using adaptive finite elements
Diamond	Alignment tool for aligning short DNA sequencing reads to a protein reference database
Ea-utils	Command-line tools for processing biological sequencing data
Eigen	high-level C++ library of template headers
Exonerate	Generic tool for sequence comparison
Fastme	Algorithms to infer phylogenies
FastQC	Quality control tool for high throughput sequence data
Fasttree	For large alignments
FastX	Command line tools for Short-Reads FASTA/FASTQ files preprocessing
FFmpeg	A cross-platform solution to record, convert and stream audio and video
FFTW	Fast fourier transforms. Ocelote has multiple versions

Fiji	Fiji is an image processing package—a “batteries-included” distribution of ImageJ2, bundling a lot of plugins which facilitate scientific image analysis.
FragGeneScan	Finds fragmented genes in short reads and predicts prokaryotic genes
FreeBayes	Bayesian genetic variant detector designed to find small polymorphisms
FreeC	Quantifying transcripts
Freesurfer	Analysis of neuroimaging data
GATK	Identifying SNPs and indels in germline DNA from Broad Institute
Gaussian	Electronic structure program. Licensed for general use
GDAL	a computer software library for reading and writing raster and vector geospatial data formats
GEOS	Geometry Engine Open Source
Git	Version control system (VCS) for tracking changes in computer files and coordinating work on those files among multiple people.
Gmap	Cross platform, open source .NET control
Gnuplot	Generates two- and three-dimensional plots of functions
Go	Programming language from Google
Gossamer	Bioinformatics suite
Gotcha	A library that provides function wrapping, interposing a wrapper function between a function and its callsites.
Grace	Grace is a Motif application for two-dimensional data visualization. Grace can transform the data using free equations, FFT, cross- and auto-correlation
Graphviz	Open source graph visualization software (included in Anaconda also)
Gromacs	Molecular dynamics software primarily designed for biomolecular systems
GSL	The GNU Scientific Library (GSL) is a numerical library for C and C++ programmers
Gurobi	Mathematical problem solver for prescriptive analytics
HDF5	Data model, library, and file format for storing and managing data
HISAT2	A fast and sensitive alignment program for mapping next-generation sequencing reads. Replaces TopHat
HMMER	Biosequence analysis using profile hidden Markov models
hpctoolkit hpctraceviewer hpcviewer	Tools for measurement and analysis of program performance
htslib	Unified C library for accessing common file formats. Also part of samtools
hwloc	Gathers information about parallel computing platforms so as to exploit them efficiently
HYPRE	Library of linear solvers featuring parallel multigrid
IDL	Restricted: <b>Interactive Data Language</b> , is a programming language used for data analysis, particularly in astronomy
Intel Compilers	Licensed compilers
Intel MPI	Intel MPI (integrated from 2019 on)
Intel Toolkit	Intel DAAL, GDB, IPP, MKL, TBB, Intel-Cluster (integrated from 2019 on)
IQ-TREE	Efficient phylogenetic software
iRods	Client - open source data management
Jags	Analysis of Bayesian hierarchical models using Markov Chain Monte Carlo simulation
Java	Programming language
Jellyfish	Fast, memory-efficient counting of k-mers in DNA. Used by Trinity

Julia	High-level, high-performance dynamic programming language for technical computing
Jupyter	Jupyter notebooks are available at the <a href="#">web service OnDemand</a>
Kallisto	Quantifying abundances of transcripts from RNA-Seq data
Lammps	Classical molecular dynamics code
Lapack	Numerical linear algebra
Libmesh	The libMesh library provides a framework for the numerical simulation of partial differential equations using arbitrary unstructured discretizations
Libpng	libpng is the official PNG reference library. It supports almost all PNG features, is extensible, and has been extensively tested
Libtool	Generic library building tool
LS-OPT (Isopt)	Design optimization
MAFFT	Multiple sequence alignment program
Maker	Portable and easily configurable genome annotation pipeline
Mathematica	Licensed: A single integrated, continually expanding system that covers the breadth and depth of technical computing
MATLAB	High-level language and interactive environment, performs computationally intensive tasks
Maven	A build automation tool used primarily for Java projects
MCL	A cluster algorithm for graphs
Meme	Suite of motif-based sequence analysis tools
MetaPhlAn	A computational tool for profiling the composition of microbial communities
Midnightcommander	A visual file manager
Migrate	Software that estimates population parameters, effective population sizes and migration rates of n populations, using genetic data
Moose	An open-source parallel finite element framework
Mothur	Software for microbial biology
Mrbayes	Provides bayesian estimation of phylogeny
MPICH/2	Freely available, portable implementation of MPI. Renamed to MPICH
MVAPICH	Library exploiting novel features and mechanisms of high-performance networking technologies
Mummer	Rapid whole genome alignment
NAMD	Molecular dynamics. Cuda version is "namd-cuda"
NCBI-vdb	A collection of tools and libraries for using data in the INSDC Sequence Read Archives.
NCDU	NCurses Disk usage analyzer
NCL (ncl-ncarg)	NCAR Command Language
NCO	Toolkit to manipulate and analyze data stored in netCDF format
NetCDF	Software libraries and self-describing, machine-independent data formats supporting the creation, access, and sharing of array-oriented scientific data
Netlogo	A programmable modeling environment for simulating natural and social phenomena
NGS-SDK	A new, domain-specific API for accessing reads, alignments and pileups produced from Next Generation Sequencing. Works with HISAT2 and SRAToolkit
OHPC	OpenHPC: Provides a variety of common, pre-built ingredients required to deploy and manage an HPC Linux cluster
OligoArray Aux	A subset of the UNAFold package for use with OligoArray
Openblas	An optimized BLAS library
OpenFOAM	Computational Fluid Dynamics software

OpenMPI	High performance message passing library
OrthoFinder	Accurate inference of orthogroups, orthologues, gene trees and rooted species
Ovito	A visualization and analysis software for output data generated in molecular dynamics, atomistic Monte-Carlo and other particle-based simulations.
Pandaseq	A program to align illumina reads
Papi	Performance application programming interface
Parallel	GNU Parallel is a shell tool for executing jobs in parallel
Paraview	ParaView is an open-source, multi-platform data analysis and visualization application.
Parflow	A parallel integrated hydrology model
ParMETIS	An MPI-based parallel library Integrated in Petsc
Pasta	Practical Alignment using Sate and TrAnsitivity. This is installed to python/2. module load python/2. Then pip list will display available modules
PCRE	A set of functions that implement regular expression pattern matching using the same syntax and semantics as Perl 5.
Peridigm	Computational peridynamics code from Sandia NL
Perl	Programming language
PETSc	A suite of data structures and routines developed by Argonne National Laboratory for the scalable (parallel) solution of scientific applications modeled by partial differential equations. The default version is built with GCC and real arithmetic. There are other modules built with Intel compilers and complex arithmetic support.
PGI	PGI Compilers and Tools
phdf	A file format, library, and utility programs for efficiently managing large and complex datasets stored in files.
phenix	A comprehensive software package for macromolecular structure determination using crystallographic (X-ray, neutron and electron) and electron cryo-microscopy
Photoscan-pro	Performs photogrammetric processing of digital images and generates 3d spatial data
Picard	Command line tools for manipulating high-throughput sequencing
pkg-config	A computer program that defines and supports a unified interface for querying installed libraries for the purpose of compiling software that depends on them
Plasma	Parallel Linear Algebra Software for Multicore Architectures
Plink	Whole genome association analysis toolset
pmix	A means of exchanging wireup information needed for interprocess communication
pnetcdf	A high-performance parallel I/O library for accessing Unidata's NetCDF, files in classic formats, specifically the formats of CDF-1, 2, and 5.
Proj	Cartographic projections and coordinate transformations library
Python	Object-oriented programming language We encourage <a href="#">the use of virtualenv</a> to build your own environment.
Qctool	Software to filter out samples or variants
Qiime	Quantitative Insights Into Microbial Ecology. Qiime2 is a package in Python3
Quantum-espresso	Materials modeling
R	Language and environment for statistical computing and graphs
RStudio	RStudio is an IDE for R, available at <a href="#">OnDemand web services</a>
RAxML	A program for sequential and parallel Maximum Likelihood-based inference of large phylogenetic trees
Relion	Program for Maximum A Posteriori refinement in cryo-electron microscopy
Remora	Resource Monitoring for Remote Applications
RepeatMasker	A program that screens DNA sequences for interspersed repeats and low complexity DNA sequences

Rings	Rigorous Investigation of Networks Generated using Simulations" is a scientific code developed in Fortran90/MPI to analyze the results of molecular dynamics simulations
Rmblast	A RepeatMasker compatible version of the standard NCBI BLAST suite. The primary difference between this distribution and the NCBI distribution is the addition of a new program "rmblastn" for use with RepeatMasker and RepeatModeler
Root-Cern	A modular scientific software framework
Ruby	A dynamic, reflective, object-oriented, general-purpose programming language.
Sagemath	A free open-source mathematics software system
Salmon	A quasi-mapping bioinformatics tool
SAMTools	Utilities for manipulating alignments in SAM format
SAS	Software suite developed by SAS Institute for advanced analytics, multivariate analyses, business intelligence, data management, and predictive analytics.
SBT	Scala build tool
Scala	General-purpose programming language
Schrodinger	Licensed: Molecular modeling and materials science.
Seqlogo	Package that takes the position weight matrix of a DNA sequence motif and plots the corresponding sequence logo.
Shapeit	Estimation of phasing for SNP sequencing data
Signalp	Package that predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms: Gram-positive prokaryotes, Gram-negative prokaryotes, and eukaryotes.
Silo	A mesh and field I/O library and scientific database
Singularity	Singularity containers let users run applications in a Linux environment of their choosing. See <a href="#">tutorial information</a>
Slim	An evolutionary simulation software package used for research and teaching
SNAP	Used by Maker
SOAPdeno vo2	A novel short-read assembly method
SPAdes	St. Petersburg genome assembler, for both standard isolates and single-cell MDA bacteria assemblies
Sparsehash	An extremely memory-efficient hash_map implementation
Spark	From Apache. Open-source distributed general-purpose cluster computing framework
Spectra	C++ library for large scale eigenvalue problems
Speedseq	An open-source genome analysis platform for rapid genome analysis and interpretation
Spparks	Kinetic Monte Carlo simulator from Sandia
SRAtoolkit	Enables reading of sequencing files from the SRA database. From NCBI
Stacks	Software pipeline for building loci from short-read sequences
Star	RNA-seq Aligner
Starfusion	Uses the STAR aligner to identify candidate fusion transcripts
Tmhmm	Package that predicts transmembrane helices in proteins.
TopHat	Fast splice junction mapper for RNA-Seq reads
TRF	Telomere Restriction Fragment (TRF) Analysis
Trilinos	Set of solvers from Sandia National Labs
Trimmomatic	A flexible trimmer for illumina sequence data
Trinity	Package which enables the efficient and robust <i>de novo</i> reconstruction of transcriptomes from RNA-Seq data
Trinotate	Annotation suite designed for automatic functional annotation of transcriptomes, particularly <i>de novo</i> assembled transcriptomes

VASP	Atomic scale materials modelling, e.g. electronic structure calculations and quantum-mechanical molecular dynamics, from first principles. <b>Special case:</b> restricted to licensed users
VCFtools	Tool providing easily accessible methods for working with complex genetic variation data in the form of VCF files
Velvet	<i>De novo</i> genomic assembler specially designed for short read sequencing technologies
Visit	Interactive parallel <b>visualization</b> and graphical analysis tool for viewing scientific data
Vtune	Intel performance profiler (vtune_amplifier_xe)
WGS	Whole Genome Shotgun Assembler for the reconstruction of genomic DNA sequence from WGS sequencing data
Wham	Whole genome Alignment Metrics
Wien2K	Software for electronic structure calculations using DFT
WRF WPS	Weather Research and Forecasting Model. <b>Special case:</b> available to Hydrology and Atmospheric Sciences
XDMF	eXtensible Data Model and Format
Xz	Free general-purpose data compression software with a high compression ratio.
zlib	A software library used for data compression.

\*\* installed on the operating system of each node. "module load xx" is not necessary.