

# Software

## Software

The following is a categorized list of software available through the UT GSAF and/or TACC's life sciences group. Each page lists a summary of the software, the hardware it is currently installed on, links to user documentation, and helpful tips.

Many of these are available via the TACC module system (use `module keyword` or `module spider` to search).

The current listing of modules and versions available on Lonestar is also posted [here](#).

In addition, the BioITeam maintains executables in `/corral-repl/utexas/BioITeam/bin`; we advise adding this directory to your path.

## Contents

- [Getting an account of GSAF's Galaxy instance](#) NOTE: Galaxy at this site is no longer actively supported.
- [Getting an account on GSAF server-fourierseq](#)
- [Getting started with Unix and Perl](#)
- [Tips for working with TACC resources](#)
- [TACC Lonestar workflow scripts](#)
- [References \(by organism\)](#)
- [General purpose tools](#)
- [Galaxy Workflows](#)
- [Microarray data analysis tools](#)
- [NGS Data Quality Control Tools](#)
- [Mappers / Aligners](#)
- [Gene prediction tools](#)
- [SNP discovery](#)
- [RNA-Seq analysis](#)
- [Splice Junction discovery](#)
- [Genome Alignment and Visualization](#)
- [De novo assembly](#)
- [Transcriptome De novo assembly](#)
- [454 Analysis tools](#)
- [Useful scripts](#)
- [Software users group meetings](#)
- [Variant calling](#)
- [ZOHO Information](#)
- [Extracting barcode split data from SOLiD 5500 XSQ files](#)
- [R and R packages](#)

## Galaxy Workflows

- [RNA-seq workflow](#)
- [BWA-SAMTOOLS workflow](#)

## General purpose tools

- [Blast](#)
- [Bioconductor](#)
- [Bioperl](#)
- [BioMart Perl APIs](#)
- [NCBI Eutils](#)
- [R](#)
- [Python Library](#)
- [Graphics programs](#)
- [BOOST libraries](#)
- [Phred, Phrap, Consed, cross\\_match, daev](#)
- [Picard](#)
- [Hmmer](#)
- [Data compression programs](#)
- [Clustering programs - MCL and usearch, uclust](#)
- [ssh - generating keys](#)

## Microarray data analysis tools

- [MeV](#)

## NGS Data Quality Control Tools

- [Fastx Toolkit](#)
- [FastQC](#)

- [Get FASTQ Format](#)

### Mappers/Aligners

- [mapreads](#) SOLiD data only, ungapped alignment
- [MAQ](#) - best for short-read SNP calling; ungapped alignment
- [muscle](#) - "old school" aligner - good for 454 amplicons
- [SOAP](#) - very fast and versatile: any read length, gapped, paired-end, SNP calling
- [SSAHA & SSAHA2](#) - like Maq, fast for ungapped mapping - SNP calling, contig placement to reference, etc.
- [Bowtie](#) - very fast, ungapped alignment. Does not support color space data
- [SHRiMP](#) - A sensitive and accurate mapper. Supports color space data and gapped alignment.
- [BFAST](#) - BLAT-like short read mapper. Natively supports SOLiD colorspace short reads.
- [BWA](#) - The successor to MAQ; a BW mapper, but which allows for gaps and handles colorspace natively.
- [GMAP and GSNAP](#) - Mappers for cDNA and very sensitive detection of short indels.
- [Mosaik](#) - A suite of alignment and reference-guided assembly tools.
- See [Category:Mapper](#) for more details.

### Gene prediction tools

- [GeneWise](#)
- [Geneid](#)
- [glimmer3](#)

### SNP discovery and Annotation

- [Corona-Lite](#) - SOLiD data only
- [Breakdancer](#)
- [Genome Analysis Tool Kit \(GATK\)](#)
- [MAQ](#) - best for short-read SNP calling; ungapped alignment
- [Picard](#)
- [SOAP](#) - very versatile: any read length, gapped, paired-end, SNP calling
- [SAMTOOLS](#)
- [Annovar](#)

### RNA-Seq Analysis

- [Tophat-Cufflinks-Cuffdiff](#), ignoring novel transcripts
- [Tophat-Cufflinks-Cuffdiff](#), allowing for novel transcripts
- [Removing duplicates from alignment output](#)

### Splice Junction discovery

---

- [Tophat- Cufflinks](#)

### Genome Alignment and Visualization

---

- [Circos](#)
- [IGV](#)
- [MaqView](#)
- [Mauve](#)
- [Affymetrix Integrated Genome Browser](#): easy to install genome browser. [Download here](#)
- [SAMStat](#)

### De novo assembly

- [MIRA](#)
- [Velvet](#)
- [ABYSS](#)
- [Using consed](#), including editing Mira assemblies
- [ABI's SOLiD de novo pipeline](#)
- [Phred, Phrap, Consed, cross\\_match, daev](#)
- [Allpaths-LG](#)
- [Mosaik](#) - A suite of alignment and reference-guided assembly tools.
- [\(Newbler, the Roche/454 assembler, is under 454 Analysis tools\)](#)

### Transcriptome de novo assembly

- [Trinity](#)
- [SOAPtrans](#)

### ABI pipelines

- [Small RNA Pipeline](#)

### 454 Analysis tools

Current Roche/454 software versions on Fouriersseq are all 2.5.3. Tarballs of various 454 software versions are available at /home/daras/454sw\*

- [Sff file manipulation tools](#) - Utilities to convert and manipulate 454 sff files.
- [GS De novo assembler](#) - Performs assembly of reads and generates contigs.
- [GS Reference mapper](#) - Maps reads to a reference genome and reports consensus and variants.
- [GS Amplicon variant analyzer](#) - For detection of variants in amplicon libraries : a small region of interest at very large coverage.
- [GS Run processor and run browser](#) - Generally run already by the GSAF, but you might want to re-process image data sometimes.
- [Georgiou Lab Amplicon scripts](#) - Matlab scripts...
- [BLAST tools](#) - Scripts for quick and dirty blasts of 454 reads and contigs to see what's going on at a global level

### Useful scripts

- [Convert ABI SOLiD data to fasta fastq](#)
- [General parser scripts](#) - scripts for parsing and filtering of fasta, fastq files and output files from different mappers; base space, color space conversion scripts.
- [Small rna analysis](#)
- [Generation of wig files from mapreads output](#)
- [Conversion of mapreads output to GFF, SAM, or BAM format](#) - These utilities can be used to convert mapreads mapping output to base space format
- [Generation of gene counts from results of mapping to genome](#) - These scripts can be used to identify the reads that correspond to genes, after mapping to the genome.
- [Get Tm \(melting temperature\), length, and %GC from a bunch of sequences](#)
- [Conversion of gene ID's from one form to another \(i.e. NCBI to Ensembl & vice-versa\)](#)
- [Quick tips on GO analysis](#)
- [Median polish to consolidate quantitations](#)
- [Make a quick venn diagram based on lists in 3 files](#)
- [Plot a read length histogram based on sequences in a fasta file](#)
- [Reverse complement for fasta files](#)
- [Tricks to preprocess SOLiD and 454 data](#)
- [Convert BLAST results to GFF](#)

### Software users group meetings

- [Small-rna data analysis](#) - Lessons learned during Sullivan data analysis
- [Mapping of short reads](#) - Comparison of few publicly available mapping tools