

■ module 経由で使えるもの

事前に設定ファイルを読み込む必要があります

```
source /apl/bio/etc/bio.sh
```

| Analysis Type   | Application name             | Description  | installed ver.  | Official URL  |
|-----------------|------------------------------|--|-----------------|---|
| Homology search | <b>blast+</b>                | Sequence Similarity Search   | 2.16.0          | <a href="https://blast.ncbi.nlm.nih.gov/Blast.cgi">https://blast.ncbi.nlm.nih.gov/Blast.cgi</a>                                     |
|                 | <b>blat</b>                  | BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 25 bases or more   | 37              | <a href="https://genome.ucsc.edu/goldenpath/help/blatSpec.html">https://genome.ucsc.edu/goldenpath/help/blatSpec.html</a>           |
|                 | <b>Diamond</b>               | DIAMOND is a sequence aligner for protein and translated DNA searches, designed for high performance analysis of big sequence data.                              | 2.1.11          | <a href="https://github.com/bbuchfink/diamond">https://github.com/bbuchfink/diamond</a>   |
|                 | <b>fasta</b>                 | Sequence Similarity Search   | 36.3.8g         |   |
|                 | <b>HH-suite</b>              | The HH-suite is an open-source software package for sensitive protein sequence searching based on the pairwise alignment of hidden Markov models (HMMs).         | 3.3.0_SSE2      | <a href="https://github.com/soedinglab/hh-suite">https://github.com/soedinglab/hh-suite</a>   |
|                 | <b>MMseq2</b>                | ultra fast and sensitive sequence search and clustering suite  | 14-7e284        | <a href="https://github.com/soedinglab/MMseqs2">https://github.com/soedinglab/MMseqs2</a>   |
|                 | <b>vsearch</b>               | VSEARCH is an open source and free of charge multithreaded 64-bit tool for processing metagenomic nucleotide sequence data. An alternative to the USEARCH        | 2.29.0          | <a href="https://github.com/torognes/vsearch">https://github.com/torognes/vsearch</a>   |
| NGS analysis    | <b>bamtools</b>              | Bamtools is a toolkit for analyzing and managing BAM files   | 2.5.2           | <a href="https://github.com/pezmaster31/bamtools">https://github.com/pezmaster31/bamtools</a>                                       |
|                 | <b>bedops</b>                | BEDOPS: the fast, highly scalable and easily-parallelizable genome analysis toolkit  | 2.4.41          | <a href="https://bedops.readthedocs.io/en/latest/index.html">https://bedops.readthedocs.io/en/latest/index.html</a>                 |
|                 | <b>BEDtools</b>              | Bedtools utilities are a tools for a wide-range of genomics analysis tasks   | 2.31.1          | <a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>                                 |
|                 | <b>Bowtie</b>                | Bowtie is an ultrafast, memory-efficient short read aligner  | 1.2.3           | <a href="https://bowtie-bio.sourceforge.net/index.shtml">https://bowtie-bio.sourceforge.net/index.shtml</a>                         |
|                 | <b>Bowtie2</b>               | Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.  | 2.5.3           | <a href="https://bowtie-bio.sourceforge.net/bowtie2/index.shtml">https://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>         |
|                 | <b>bwa</b><br><b>bwa-mem</b> | Burrows-Wheeler Aligner (BWA) is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence                        | 0.17.7<br>2.2.1 | <a href="https://github.com/lh3/bwa">https://github.com/lh3/bwa</a>   |
|                 | <b>Cufflinks</b>             | Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples                             |                 | <a href="https://github.com/cole-trapnell-lab/cufflinks">https://github.com/cole-trapnell-lab/cufflinks</a>                         |
|                 | <b>fastQC</b>                | A quality control tool for high throughput sequence data.  |                 | <a href="https://www.bioinformatics.babraham.ac.uk/projects/fastqc/">https://www.bioinformatics.babraham.ac.uk/projects/fastqc/</a> |
|                 | <b>fastp</b>                 | A quality control tool for high throughput sequence data.  | 0.23.4          | <a href="https://github.com/OpenGene/fastp">https://github.com/OpenGene/fastp</a>   |
|                 | <b>GATK</b>                  | The GATK is the industry standard for identifying SNPs and indels in germline DNA and RNAseq data.   | 4.3.0.0         | <a href="https://github.com/broadinstitute/gatk">https://github.com/broadinstitute/gatk</a>   |
|                 | <b>hisat2</b>                | HISAT2 is a fast and sensitive alignment program for mapping next-generation sequencing reads (both DNA and RNA).  | 2.2.1           | <a href="https://daehwankimlab.github.io/hisat2/">https://daehwankimlab.github.io/hisat2/</a>                                       |
|                 | <b>kallisto</b>              | kallisto is a program for quantifying abundances of transcripts from RNA-Seq data, or more generally of target sequences using high-throughput sequencing reads. | 0.46.2          | <a href="https://github.com/pachterlab/kallisto">https://github.com/pachterlab/kallisto</a>   |
|                 | <b>MACS2</b>                 | Model-based Analysis of ChIP-Seq on short reads sequencers such as Genome Analyzer (Illumina / Solexa)   | 2.1.2           | <a href="https://github.com/macs3-project/MACS">https://github.com/macs3-project/MACS</a>   |
|                 | <b>RSEM</b>                  | RNA-Seq by Expectation-Maximization  | 1.3.3           | <a href="https://github.com/deweylab/RSEM">https://github.com/deweylab/RSEM</a>   |
|                 | <b>Salmon</b>                | Salmon is a tool for quantifying the expression of transcripts using RNA-seq data  | 1.10.0          | <a href="https://combine-lab.github.io/salmon/">https://combine-lab.github.io/salmon/</a>   |

| Analysis Type                 | Application name     | Description  | installed ver. | Official URL  |
|-------------------------------|----------------------|--|----------------|---|
|                               | <b>samtools</b>      | SAM Tools provide various utilities for manipulating alignments in the SAM format  | 1.19.2         | <a href="https://www.htslib.org/">https://www.htslib.org/</a>   |
|                               | <b>seqkit</b>        | SeqKit - a cross-platform and ultrafast toolkit for FASTA/Q file manipulation  | 2.10.0         | <a href="https://github.com/shenwei356/seqkit">https://github.com/shenwei356/seqkit</a>   |
|                               | <b>soap</b>          | Short Oligonucleotide Analysis Package   | 2.21           | <a href="https://github.com/ShujiaHuang/SOAPaligner">https://github.com/ShujiaHuang/SOAPaligner</a>   |
|                               | <b>SRAtoolkit</b>    | The SRA Toolkit and SDK from NCBI is a collection of tools and libraries for using data in the INSDC Sequence Read Archives.                   | 3.0.0          | <a href="https://github.com/ncbi/sra-tools">https://github.com/ncbi/sra-tools</a>   |
|                               | <b>STAR</b>          | Spliced Transcripts Alignment to a Reference   | 2.7.11b        | <a href="https://github.com/alexdobin/STAR">https://github.com/alexdobin/STAR</a>   |
|                               | <b>Stringtie</b>     | StringTie is a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts.   | 3.0.0          | <a href="https://ccb.jhu.edu/software/stringtie/">https://ccb.jhu.edu/software/stringtie/</a>   |
|                               | <b>Tophat</b>        | TopHat is a fast splice junction mapper for RNA-Seq reads  | 2.1.1          | <a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>   |
| Genome (transcript) Assembler | <b>ABYSS</b>         | Assembly By Short Sequences - a de novo, parallel, paired-end sequence assembler   | 2.3.4          | <a href="https://github.com/bcgsc/abyss">https://github.com/bcgsc/abyss</a>   |
|                               | <b>Allpaths-LG</b>   | The new short read genome assembler.   | 52488          | <a href="https://www.rcac.purdue.edu/software/allpathslg">https://www.rcac.purdue.edu/software/allpathslg</a>   |
|                               | <b>canu</b>          | Canu is a fork of the Celera Assembler designed for high-noise single-molecule sequencing (such as the PacBio RSII or Oxford Nanopore MinION). | 2.2            | <a href="https://github.com/marbl/canu">https://github.com/marbl/canu</a>   |
|                               | <b>hifiasm</b>       | Hifiasm is a fast haplotype-resolved de novo assembler for PacBio HiFi reads.  | 0.18.5         | <a href="https://github.com/chhylp123/hifiasm">https://github.com/chhylp123/hifiasm</a>   |
|                               | <b>MaSuRCA</b>       | MaSuRCA (Maryland Super-Read Celera Assembler) genome assembly software  | 4.0.7          | <a href="https://github.com/alekseyzimin/masurca">https://github.com/alekseyzimin/masurca</a>   |
|                               | <b>NECAT</b>         | NECAT is an error correction and de-novo assembly tool for Nanopore long noisy reads.  | 0.0.1          | <a href="https://github.com/xiaochuanle/NECAT">https://github.com/xiaochuanle/NECAT</a>   |
|                               | <b>SPAdes</b>        | SPAdes St. Petersburg genome assembler: SPAdes is an assembly toolkit containing various assembly pipelines.                                   | 4.1.0          | <a href="https://github.com/ablab/spades">https://github.com/ablab/spades</a>   |
|                               | <b>Trinityrnaseq</b> | Novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-seq data   | 2.15.1         | <a href="https://github.com/trinityrnaseq/trinityrnaseq">https://github.com/trinityrnaseq/trinityrnaseq</a>   |
|                               | <b>velvet</b>        | Sequence assembler for very short reads  | 1.2.10         | <a href="https://github.com/dzerbino/velvet">https://github.com/dzerbino/velvet</a>   |
|                               | <b>soap denovo</b>   | Short Oligonucleotide Analysis Package genome assembler  | 2.04           | <a href="https://github.com/ShujiaHuang/SOAPaligner">https://github.com/ShujiaHuang/SOAPaligner</a>   |
| Pairwise Alignment            | <b>wgs</b>           | same as Celera Assembler: Whole genome assembler   | 8.2            | <a href="https://github.com/alekseyzimin/wgs">https://github.com/alekseyzimin/wgs</a>   |
|                               | <b>lastz</b>         | A tool for (1) aligning two DNA sequences, and (2) inferring appropriate scoring parameters automatically                                      | 1.04           | <a href="https://github.com/lastz/lastz">https://github.com/lastz/lastz</a>   |
| Multiple Alignment            | <b>MUMmer</b>        | MUMmer is a system for rapidly aligning entire genomes, whether in complete or draft form.   | 4.0.0          | <a href="https://github.com/mummer4/mummer">https://github.com/mummer4/mummer</a>   |
|                               | <b>clustal Omega</b> | Fast, accurate, scalable multiple sequence alignment for proteins  | 1.2.4          | <a href="http://www.clustal.org/omega/">http://www.clustal.org/omega/</a>   |
|                               | <b>clustalw</b>      | Multiple Sequence Alignment  | 1.83           | <a href="http://www.clustal.org/clustal2/">http://www.clustal.org/clustal2/</a>   |
|                               | <b>clustalw2</b>     | Multiple Sequence Alignment  | 2.1            | <a href="http://www.clustal.org/clustal2/">http://www.clustal.org/clustal2/</a>   |
|                               | <b>FAMSA</b>         | Progressive algorithm for large-scale multiple sequence alignments.  | 1.6.2          | <a href="https://github.com/refresh-bio/FAMSA">https://github.com/refresh-bio/FAMSA</a>   |
|                               | <b>Gblocks</b>       | Gblocks eliminates poorly aligned positions and divergent regions of an alignment of DNA or protein sequences                                  | 0.91b          | <a href="https://www.biologiaevolutiva.org/jcastresana/Gblocks.html">https://www.biologiaevolutiva.org/jcastresana/Gblocks.html</a>   |
|                               | <b>MAFFT</b>         | MAFFT is a multiple sequence alignment program   | 7.526          | <a href="https://mafft.cbrc.jp/alignment/software/">https://mafft.cbrc.jp/alignment/software/</a>   |
|                               | <b>muscle</b>        | Multiple Sequence Alignment faster and more accurate than clustalw   | 5.1            | <a href="https://www.drive5.com/muscle/">https://www.drive5.com/muscle/</a>   |
| Database search               | <b>t_coffee</b>      | Multiple sequence alignment package  | 12.00.7        | <a href="https://tcoffee.org/Projects/tcoffee/index.html">https://tcoffee.org/Projects/tcoffee/index.html</a>   |
|                               | <b>dbget</b>         | DBGET is an integrated database retrieval system for major biological databases  |                | <a href="https://www.kanehisa.jp/ja/about_dbget_ja.html">https://www.kanehisa.jp/ja/about_dbget_ja.html</a>   |
|                               | <b>CAP3</b>          | Multiple Sequence Alignment  | 122107         | <a href="https://faculty.sites.iastate.edu/xqhuang/cap3-and-pcap-sequence-and-genome-assembly-programs">https://faculty.sites.iastate.edu/xqhuang/cap3-and-pcap-sequence-and-genome-assembly-programs</a> |
|                               | <b>consed</b>        | Assembly Editor  | 29.0           | <a href="http://bozeman.mbt.washington.edu/consed/consed.html">http://bozeman.mbt.washington.edu/consed/consed.html</a>   |
|                               | <b>Phrap</b>         | Phrap is a program for assembling shotgun DNA sequence data  | 1.090518       | <a href="http://www.phrap.org/phredphrapconsed.html">http://www.phrap.org/phredphrapconsed.html</a>   |

| Analysis Type              | Application name        | Description   | installed ver.     | Official URL  |
|----------------------------|-------------------------|---|--------------------|---|
| Genome Assembler           | <b>Phred</b>            | The phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base   | 071220             | <a href="http://www.phrap.org/phredphrapconsed.html">http://www.phrap.org/phredphrapconsed.html</a>   |
|                            | <b>TGICL</b>            | Multiple Sequence Alignment (for huge data set)   | 2.1                | it's gone : <a href="http://compbio.dfci.harvard.edu/tgi/software/">http://compbio.dfci.harvard.edu/tgi/software/</a>                               |
| Gene prediction            | <b>Augustus</b>         | AUGUSTUS is a program that predicts genes in eukaryotic genomic sequences   | 3.5.0              | <a href="https://github.com/Gaius-Augustus/Augustus">https://github.com/Gaius-Augustus/Augustus</a>   |
|                            | <b>Genemark</b>         | A family of gene prediction programs  | 4.69               | <a href="https://genemark.bme.gatech.edu/">https://genemark.bme.gatech.edu/</a>   |
|                            | <b>genscan</b>          | Gene prediction   |                    | <a href="https://www.genes.mit.edu/GENSCAN.html">https://www.genes.mit.edu/GENSCAN.html</a>   |
|                            | <b>glimmer</b>          | Glimmer is a system for finding genes in microbial DNA  | 30.02              | <a href="https://ccb.jhu.edu/software/glimmerhmm/">https://ccb.jhu.edu/software/glimmerhmm/</a>   |
|                            | <b>glimmerhmm</b>       |   | 3.0.4              |   |
|                            | <b>TSEBRA</b>           | TSEBRA is a combiner tool that selects transcripts from gene predictions based on the support by extrinsic evidence in form of introns and start/stop codons  |                    | <a href="https://github.com/Gaius-Augustus/TSEBRA">https://github.com/Gaius-Augustus/TSEBRA</a>   |
|                            | <b>Metaeuk</b>          | MetaEuk - sensitive, high-throughput gene discovery and annotation for large-scale eukaryotic metagenomics  | 7-bba0d80          | <a href="https://github.com/soedinglab/metaeuk">https://github.com/soedinglab/metaeuk</a>   |
|                            | <b>Prodigal</b>         | Prodigal: Fast, reliable protein-coding gene prediction for prokaryotic genomes.  | 2.6.3              | <a href="https://github.com/hyattpd/Prodigal">https://github.com/hyattpd/Prodigal</a>   |
| Motif search               | <b>HMMER</b>            | Biosequence analysis using profile HMM  | 3.4                | <a href="http://hmmer.org/">http://hmmer.org/</a>   |
|                            | <b>Interproscan</b>     | A tool that combines different protein signature recognition methods into one resource  | 5.73-104.0         | <a href="https://www.ebi.ac.uk/interpro/">https://www.ebi.ac.uk/interpro/</a>   |
|                            | <b>meme</b>             | Multiple Em for Motif Elicitation   | 5.4.1              | <a href="https://meme-suite.org/meme/index.html">https://meme-suite.org/meme/index.html</a>   |
| Functional annotation      | <b>eggNOG-Mapper</b>    | A tool for fast functional annotation of novel sequences.   | 2.1.12             | <a href="https://github.com/eggnogdb/eggno-mapper">https://github.com/eggnogdb/eggno-mapper</a>   |
| phylogenetic tree analysis | <b>mrBayes</b>          | MrBayes is a program for Bayesian inference and model choice across a wide range of phylogenetic and evolutionary models  | 3.2.7<br>3.2.7.mpi | <a href="https://nbisweden.github.io/MrBayes/">https://nbisweden.github.io/MrBayes/</a>   |
|                            | <b>njplot</b>           | Njplot is a tree drawing program  | 2.4                | <a href="https://doua.prabi.fr/software/njplot">https://doua.prabi.fr/software/njplot</a>   |
|                            | <b>paup</b>             | Tools for inferring and interpreting phylogenetic trees   | 4b10               | <a href="https://paup.phylosolutions.com/">https://paup.phylosolutions.com/</a>   |
|                            | <b>Phylip</b>           | A package of programs for inferring phylogenies (evolutionary trees)  | 3.697              | <a href="https://phylipweb.github.io/phylip/">https://phylipweb.github.io/phylip/</a>   |
|                            | <b>PhyML</b>            | PhyML is a phylogeny software based on the maximum-likelihood principle   | 3.3                | <a href="https://github.com/stephaneguindon/phyml">https://github.com/stephaneguindon/phyml</a>   |
|                            | <b>RAxML (raxmlHPC)</b> | RAxML - Randomized Axelerated Maximum Likelihood  | 8.2.13             | <a href="https://github.com/stamatak/standard-RAxML/">https://github.com/stamatak/standard-RAxML/</a>   |
|                            | <b>tree-puzzle</b>      | Program to reconstruct phylogenetic trees from molecular sequence data by maximum likelihood  | 5.3.rc16           | <a href="http://www.tree-puzzle.de/">http://www.tree-puzzle.de/</a>   |
| Single cell analysis       | <b>CellRanger</b>       | A set of analysis pipelines that perform sample demultiplexing, barcode processing, single cell 3' and 5' gene counting, V(D)J transcript sequence assembly and annotation, and Feature Barcode analysis from single cell data. | 8.0.1              | <a href="https://www.10xgenomics.com/jp/support/software/cell-ranger/latest">https://www.10xgenomics.com/jp/support/software/cell-ranger/latest</a> |
|                            | <b>MARVEL</b>           | MARVEL is an R package developed for alternative splicing analysis at single-cell resolution.   | 1.0.0              | <a href="https://github.com/wenweixiong/MARVEL">https://github.com/wenweixiong/MARVEL</a>   |
| Repeat Masking             | <b>RepeatMasker</b>     | RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences  | 4.1.8              | <a href="https://www.repeatmasker.org/">https://www.repeatmasker.org/</a>   |
|                            | <b>RepeatModeler</b>    | RepeatModeler is a de novo transposable element (TE) family identification and modeling package.  | 2.0.6              | <a href="https://www.repeatmasker.org/RepeatModeler/">https://www.repeatmasker.org/RepeatModeler/</a>   |
| tRNA search                | <b>tRNAscan-SE</b>      | tRNAscan-SE pioneers the large-scale use of covariance models to annotate tRNA genes in genomes   | 2.0.5              | <a href="https://trna.ucsc.edu/tRNAscan-SE/">https://trna.ucsc.edu/tRNAscan-SE/</a>   |
| Other                      | <b>TransDecoder</b>     | TransDecoder identifies candidate coding regions within transcript sequences.   | 5.7.1              | <a href="https://github.com/TransDecoder/TransDecoder">https://github.com/TransDecoder/TransDecoder</a>   |

■ **apptainer** で利用するためのコンテナ.sif ファイルが用意されているもの

- 作業ディレクトリにシンボリックリンクを作ってください
- パス：

|                      |   | Version | .sif file                                   | remarks                                    | Official URL  |
|----------------------|---|---------|---|--|---|
| <b>AGAT</b>          | AGAT has the power to check, fix, pad missing information (features/attributes) of any kind of GTF and GFF to create complete, sorted and standardised gff3 format.   | 1.4.1   | AGAT/1.4.1/agat_1.4.1--pl5321hdfd78af_0.sif |  | <a href="https://github.com/NBISweden/AGAT">https://github.com/NBISweden/AGAT</a>               |
| <b>BRAKER</b>        | BRAKER is a program that predicts genes in eukaryotic genomic sequences   | 3.0.2   | BRAKER/3.0.2/braker3.sif                    |  | <a href="https://github.com/Gaius-Augustus/BRAKER">https://github.com/Gaius-Augustus/BRAKER</a> |
| <b>BUSCO</b>         | Assessing genome assembly and annotation completeness with single-copy orthologs  | 5.8.0   | BUSCO/5.8.0/busco580.sif                    | Required Options :<br>--<br><b>offline</b> | <a href="https://busco.ezlab.org/">https://busco.ezlab.org/</a>                                 |
| <b>DeepConsensus</b> | DeepConsensus uses gap-aware sequence transformers to correct errors in Pacific Biosciences (PacBio) Circular Consensus Sequencing (CCS) data.                        | 1.2.0   | DeepConsensus/1.2.0/deepconsensus.sif       |  | <a href="https://github.com/google/deepconsensus">https://github.com/google/deepconsensus</a>   |
| <b>DeepTMHMM</b>     | DeepTMHMM is currently the most complete and best-performing method for the prediction of the topology of both alpha-helical and beta-barrel transmembrane proteins.  | 1.0.42  | DeepTMHMM/1.0.42/deeptmhmm_edit_g.sif       |  | <a href="https://dtu.biolib.com/DeepTMHMM">https://dtu.biolib.com/DeepTMHMM</a>                 |
| <b>EpiTyping</b>     | EpiTyping is a tool for detecting imprinting and X-chromosome inactivation status from RNA-seq  | 1       | EpiTyping/epityping.sif                     |  | <a href="https://github.com/Gal-Keshet/EpiTyping">https://github.com/Gal-Keshet/EpiTyping</a>   |
| <b>GALBA</b>         | GALBA uses the protein sequences of several (few) or one closely related species to generate a training gene set for AUGUSTUS with either miniprot or GenomeThreader. | 1.0.7   | GALBA/1.0.7/galba107_aug35.sif              | with Augustus 3.5.0                        | <a href="https://github.com/Gaius-Augustus/GALBA">https://github.com/Gaius-Augustus/GALBA</a>   |
| <b>GATK</b>          | The GATK is the industry standard for identifying SNPs and indels in germline DNA and RNAseq data.  | 4.0.1   | GATK/4.0.1/gatk-4.sif                       |  | <a href="https://github.com/broadinstitute/gatk">https://github.com/broadinstitute/gatk</a>     |
| <b>ipyrad</b>        | An interactive assembly and analysis toolkit for restriction-site associated DNA (RAD-seq) and related data types.  | 0.9.81  | ipyrad/0.9.81/ipyrad_0.9.81--pyh5e36f6f_0   |  | <a href="https://ipyrad.readthedocs.io/en/master/">https://ipyrad.readthedocs.io/en/master/</a> |

|                      |   |       |  |  |   |
|----------------------|---|-------|--|--|---|
| <b>PASA Pipeline</b> | PASA, acronym for Program to Assemble Spliced Alignments (and pronounced 'pass-uh'), is a eukaryotic genome annotation tool that exploits spliced alignments of expressed transcript sequences to automatically model gene structures | 2.5.3 | /apl/bio/container/PASA/2.5.3/pasapipeline.v2.5.3.simg |  | <a href="https://github.com/PASApipeline/PASApipeline">https://github.com/PASApipeline/PASApipeline</a> |
|----------------------|---|-------|--|--|---|