パッケージプログラム一覧（基礎生物学）

■ 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	**blast+**	Sequence Similarity Search	2.16.0	https://blast.ncbi.nlm.nih.gov/Blast.cgi
	blat	BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 25 bases or more	37	https://genome.ucsc.edu/goldenpath/help/blatSpec.html
	Diamond	DIAMOND is a sequence aligner for protein and translated DNA searches, designed for high performance analysis of big sequence data.	2.1.16	https://github.com/bbuchfink/diamond
	fasta	Sequence Similarity Search	36.3.8g	
	HH-suite	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	https://github.com/soedinglab/hh-suite
	MMseq2	ultra fast and sensitive sequence search and clustering suite	18-8cc5c	https://github.com/soedinglab/MMseqs2
	vsearch	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	https://github.com/torognes/vsearch
NGS analysis	**bamtools**	Bamtools is a toolkit for analyzing and managing BAM files	2.5.2	https://github.com/pezmaster31/bamtools
	bedops	BEDOPS: the fast, highly scalable and easily-parallelizable genome analysis toolkit	2.4.41	https://bedops.readthedocs.io/en/latest/index.html
	BEDtools	Bedtools utilities are a tools for a wide-range of genomics analysis tasks	2.31.1	https://bedtools.readthedocs.io/en/latest/
	Bowtie	Bowtie is an ultrafast, memory-efficient short read aligner	1.2.3	https://bowtie-bio.sourceforge.net/index.shtml
	Bowtie2	Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.	2.5.3	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
	bwa **bwa-mem**	Burrows-Wheeler Aligner (BWA) is an efficient program that aligns relatively short nucleotide sequences against a long reference sequence	0.17.7 2.2.1	https://github.com/lh3/bwa
	Cufflinks	Cufflinks assembles transcripts, estimates their abundances, and tests for differential expression and regulation in RNA-Seq samples		https://github.com/cole-trapnell-lab/cufflinks
	fastQC	A quality control tool for high throughput sequence data.		https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
	fastp	A quality control tool for high throughput sequence data.	1.0.1	https://github.com/OpenGene/fastp
	hisat2	HISAT2 is a fast and sensitive alignment program for mapping next-generation sequencing reads (both DNA and RNA).	2.2.1	https://daehwankimlab.github.io/hisat2/
	kallisto	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	https://github.com/pachterlab/kallisto
	MACS2	Model-based Analysis of ChIP-Seq on short reads sequencers such as Genome Analyzer (Illumina / Solexa)	2.1.2	https://github.com/macs3-project/MACS
	RSEM	RNA-Seq by Expectation-Maximization	1.3.3	https://github.com/deweylab/RSEM
	Salmon	Salmon is a tool for quantifying the expression of transcripts using RNA-seq data	1.10.0	https://combine-lab.github.io/salmon/
	samtools	SAM Tools provide various utilities for manipulating alignments in the SAM format	1.19.2	https://www.htslib.org/

Analysis Type	Application name	Description	installed ver.	Official URL
	seqkit	SeqKit - a cross-platform and ultrafast toolkit for FASTA/Q file manipulation	2.10.0	https://github.com/shenwei356/seqkit
	soap	Short Oligonucleotide Analysis Package	2.21	https://github.com/ShujiaHuang/SOAPaligner
	SRAtoolkit	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	https://github.com/ncbi/sra-tools
	STAR	Spliced Transcripts Alignment to a Reference	2.7.11b	https://github.com/alexdobin/STAR
	Stringtie	StringTie is a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts.	3.0.0	https://ccb.jhu.edu/software/stringtie/
	Tophat	TopHat is a fast splice junction mapper for RNA-Seq reads	2.1.1	https://ccb.jhu.edu/software/tophat/index.shtml
Metagenome analysis	**CAT_pack**	A pipelines for the taxonomic classification of long DNA sequences and metagenome assembled genomes (MAGs / bins)	6.0.1	https://github.com/MGXlab/CAT_pack
Genome (transcript) Assembler	**ABySS**	Assembly By Short Sequences - a de novo, parallel, paired-end sequence assembler	2.3.4	https://github.com/bcgsc/abyss
	Allpaths-LG	The new short read genome assembler.	52488	https://www.rcac.purdue.edu/software/allpathslg
	canu	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	https://github.com/marbl/canu
	hifiasm	Hifiasm is a fast haplotype-resolved de novo assembler for PacBio HiFi reads.	0.18.5	https://github.com/chhylp123/hifiasm
	MaSuRCA	MaSuRCA (Maryland Super-Read Celera Assembler) genome assembly software	4.0.7	https://github.com/alekseyzimin/masurca
	NECAT	NECAT is an error correction and de-novo assembly tool for Nanopore long noisy reads.	0.0.1	https://github.com/xiaochuanle/NECAT
	SPAdes	SPAdes St. Petersburg genome assembler: SPAdes is an assembly toolkit containing various assembly pipelines.	4.2.0	https://github.com/ablab/spades
	Trinityrnaseq	Novel method for the efficient and robust de novo reconstruction of transcriptomes from RNA-seq data	2.15.1	https://github.com/trinityrnaseq/trinityrnaseq
	velvet	Sequence assembler for very short reads	1.2.10	https://github.com/dzerbino/velvet
	soap denovo	Short Oligonucleotide Analysis Package genome assembler	2.04	https://github.com/ShujiaHuang/SOAPaligner
	wgs	same as Celera Assembler: Whole genome assembler	8.2	https://github.com/alekseyzimin/wgs
Pairwise Alignment	**lastz**	A tool for (1) aligning two DNA sequences, and (2) inferring appropriate scoring parameters automatically	1.04	https://github.com/lastz/lastz
	MUMmer	MUMmer is a system for rapidly aligning entire genomes, whether in complete or draft form.	4.0.0	https://github.com/mummer4/mummer
Multiple Alignment	**clustal Omega**	Fast, accurate, scalable multiple sequence alignment for proteins	1.2.4	http://www.clustal.org/omega/
	clustalw	Multiple Sequence Alignment	1.83	http://www.clustal.org/clustal2/
	clustalw2	Multiple Sequence Alignment	2.1	http://www.clustal.org/clustal2/
	FAMSA	Progressive algorithm for large-scale multiple sequence alignments.	2.4.1	https://github.com/refresh-bio/FAMSA
	Gblocks	Gblocks eliminates poorly aligned positions and divergent regions of an alignment of DNA or protein sequences	0.91b	https://www.biologiaevolutiva.org/jcastresana/Gblocks.html
	MAFFT	MAFFT is a multiple sequence alignment program	7.526	https://mafft.cbrc.jp/alignment/software/
	muscle	Multiple Sequence Alignment faster and more accurate than clustalw	5.1	https://www.drive5.com/muscle/
	t_coffee	Multiple sequence alignment package	12.00.7	https://tcoffee.org/Projects/tcoffee/index.html
Genome Aligner	**minimap2**	A versatile sequence alignment program that aligns DNA or mRNA sequences against a large reference database	2-2.30	https://github.com/lh3/minimap2
	miniprot	Miniprot aligns a protein sequence against a genome with affine gap penalty, splicing and frameshift	0.18	https://github.com/lh3/miniprot

Analysis Type	Application name	Description	installed ver.	Official URL
Datbase search	**dbget**	DBGET is an integrated database retrieval system for major biological databases		https://www.kanehisa.jp/ja/about_dbget_ja.html
Sequence Assembler	**CAP3**	Multiple Sequence Alignment	122107	https://faculty.sites.iastate.edu/xqhuang/cap3-and-pcap-sequence-and-genome-assembly-programs
	consed	Assembly Editor	29.0	http://bozeman.mbt.washington.edu/consed/consed.html
	Phrap	Phrap is a program for assembling shotgun DNA sequence data	1.090518	http://www.phrap.org/phredphrapconsed.html
	Phred	The phred software reads DNA sequencing trace files, calls bases, and assigns a quality value to each called base	071220	http://www.phrap.org/phredphrapconsed.html
	TGICL	Multiple Sequence Alignment (for huge data set)	2.1	it's gone : http://compbio.dfci.harvard.edu/tgi/software/
Gene prediction	**Augustus**	AUGUSTUS is a program that predicts genes in eukaryotic genomic sequences	3.5.0	https://github.com/Gaius-Augustus/Augustus
	Genemark	A family of gene prediction programs	4.69	https://genemark.bme.gatech.edu/
	genscan	Gene prediction		https://www.genes.mit.edu/GENSCAN.html
	glimmer **glimmerhmm**	Glimmer is a system for finding genes in microbial DNA	30.02 3.0.4	https://ccb.jhu.edu/software/glimmerhmm/
	TSEBRA	TSEBRA is a combiner tool that selects transcripts from gene predictions based on the support by extrisic evidence in form of introns and start/stop codons		https://github.com/Gaius-Augustus/TSEBRA
	Metaeuk	MetaEuk - sensitive, high-throughput gene discovery and annotation for large-scale eukaryotic metagenomics	7-bba0d80	https://github.com/soedinglab/metaeuk
	Prodigal	Prodigal: Fast, reliable protein-coding gene prediction for prokaryotic genomes.	2.6.3	https://github.com/hyattpd/Prodigal
Motif search	**HMMER**	Biosequence analysis using profile HMM	3.4	http://hmmer.org/
	Interproscan	A tool that combines different protein signature recognition methods into one resource	5.75-106.0	https://www.ebi.ac.uk/interpro/
	meme	Multiple Em for Motif Elicitation	5.4.1	https://meme-suite.org/meme/index.html
Functional annotation	**eggNOG-Mapper**	A tool for fast functional annotation of novel sequences.	2.1.12	https://github.com/eggnogdb/eggnog-mapper
phylogenetic tree analysis	**mrbayes**	MrBayes is a program for Bayesian inference and model choice across a wide range of phylogenetic and evolutionary models	3.2.7 3.2.7.mpi	https://nbisweden.github.io/MrBayes/
	njplot	NJplot is a tree drawing program	2.4	https://doua.prabi.fr/software/njplot
	paup	Tools for inferring and interpreting phylogenetic trees	4b10	https://paup.phylosolutions.com/
	Phylip	A package of programs for inferring phylogenies (evolutionary trees)	3.697	https://phylipweb.github.io/phylip/
	PhyML	PhyML is a phylogeny software based on the maximum-likelihood principle	3.3	https://github.com/stephaneguindon/phyml
	RAxML (raxmlHPC)	RAxML - Randomized Axelerated Maximum Likelihood	8.2.13	https://github.com/stamatak/standard-RAxML/
	tree-puzzle	Program to reconstruct phylogenetic trees from molecular sequence data by maximum likelihood	5.3.rc16	http://www.tree-puzzle.de/
Single cell analysis	**CellRanger**	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	https://www.10xgenomics.com/jp/support/software/cell-ranger/latest
	MARVEL	MARVEL is an R package developed for alternative splicing analysis at single-cell resolution.	1.0.0	https://github.com/wenweixiong/MARVEL
Repeat Masking	**RepeatMasker**	RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences	4.1.8	https://www.repeatmasker.org/
	RepeatModeler	RepeatModeler is a de novo transposable element (TE) family identification and modeling package.	2.0.6	https://www.repeatmasker.org/RepeatModeler/
tRNA search	**tRNAscan-SE**	tRNAscan-SE pioneers the large-scale use of covariance models to annotate tRNA genes in genomes	2.0.5	https://trna.ucsc.edu/tRNAscan-SE/

Analysis Type	**Application name**	**Description**	**installed ver.**	**Official URL**
Other	**TransDecoder**	TransDecoder identifies candidate coding regions within transcript sequences.	5.7.1	https://github.com/TransDecoder/TransDecoder
	R		4.5.1	https://www.r-project.org/

■ 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ご利用ください
- パス：

```
/apl/bio/container/
```

		Version	.sif file	remarks	Official URL
AGAT	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		https://github.com/NBISweden/AGAT
BRAKER	BRAKER is a program that predicts genes in eukaryotic genomic sequences	3.0.2	BRAKER/3.0.2/braker3.sif		https://github.com/Gaius-Augustus/BRAKER
BUSCO	Assessing genome assembly and annotation completeness with single-copy orthologs	5.8.0	BUSCO/5.8.0/busco580.sif	Required Options : **--offline**	https://busco.ezlab.org/
DeepConsensus	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		https://github.com/google/deepconsensus
DeepTMHMM	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		https://dtu.biolib.com/DeepTMHMM
EpiTyping	EpiTyping is a tool for detecting imprinting and X-chromosome inactivation status from RNA-seq	1	EpiTyping/epityping.sif		https://github.com/Gal-Keshet/EpiTyping
GALBA	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	https://github.com/Gaius-Augustus/GALBA
GATK	The GATK is the industry standard for identifying SNPs and indels in germline DNA and RNAseq data.	4.6.2	GATK/4.6.2/gatk-426.sif		https://github.com/broadinstitute/gatk

ipyrad	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		https://ipyrad.readthedocs.io/en/master/
PASA Pipeline	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		https://github.com/PASApipeline/PASApipeline