パッケージプログラム一覧（基礎生物学）

? module ????????

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
Homology search	blast+	Sequence Similarity Search
	blat	BLAT on DNA is designed to quickly find sequences of 95% and grea
	Diamond	DIAMOND is a sequence aligner for protein and translated DNA sear
	fasta	Sequence Similarity Search
	HH-suite	The HH-suite is an open-source software package for sensitive protei
	MMseq2	ultra fast and sensitive sequence search and clustering suite
	vsearch	VSEARCH is an open source and free of charge multithreaded 64-bit
NGS analysis	bamtools	Bamtools is a toolkit for analyzing and managing BAM files
	bedops	BEDOPS: the fast, highly scalable and easily-parallelizable genome a
	BEDtools	Bedtools utilities are a tools for a wide-range of genomics analysis ta
	Bowtie	Bowtie is an ultrafast, memory-efficient short read aligner
	Bowtie2	Bowtie 2 is an ultrafast and memory-efficient tool for aligning sequen
	bwa bwa-mem	Burrows-Wheeler Aligner (BWA) is an efficient program that aligns rel
	Cufflinks	Cufflinks assembles transcripts, estimates their abundances, and test
	fastQC	A quality control tool for high throughput sequence data.
	fastp	A quality control tool for high throughput sequence data.
	hisat2	HISAT2 is a fast and sensitive alignment program for mapping next-g
	kallisto	kallisto is a program for quantifying abundances of transcripts from R
	MACS2	Model-based Analysis of ChIP-Seq on short reads sequencers such a
	RSEM	RNA-Seq by Expectation-Maximization
	Salmon	Salmon is a tool for quantifying the expression of transcripts using RN
	samtools	SAM Tools provide various utilities for manipulating alignments in the

Analysis Type	Application name	Description
seqkit	SeqKit - a cross-platform and ultrafast toolkit for FASTA/Q file manipulation	2.10.0
soap	Short Oligonucleotide Analysis Package	2.21
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
STAR	Spliced Transcripts Alignment to a Reference	2.7.11b
Stringtie	StringTie is a fast and highly efficient assembler of RNA-Seq alignments into potential transcripts.	3.0.0
Tophat	TopHat is a fast splice junction mapper for RNA-Seq reads	2.1.1
Metagenome analysis	CAT_pack	A pipelines for the taxonomic classification of long DNA sequences a
Genome (transcript) Assembler	ABySS	Assembly By Short Sequences - a de novo, parallel, paired-end seque
	Allpaths-LG	The new short read genome assembler.
	canu	Canu is a fork of the Celera Assembler designed for high-noise single
	hifiasm	Hifiasm is a fast haplotype-resolved de novo assembler for PacBio Hi
	MaSuRCA	MaSuRCA (Maryland Super-Read Celera Assembler) genome assemb
	NECAT	NECAT is an error correction and de-novo assembly tool for Nanopor
	SPAdes	SPAdes St. Petersburg genome assembler: SPAdes is an assembly to
	Trinityrnaseq	Novel method for the efficient and robust de novo reconstruction of
	velvet	Sequence assembler for very short reads
	soap denovo	Short Oligonucleotide Analysis Package genome assembler
	wgs	same as Celera Assembler: Whole genome assembler
Pairwise Alignment	lastz	A tool for (1) aligning two DNA sequences, and (2) inferring appropria
	MUMmer	MUMmer is a system for rapidly aligning entire genomes, whether in
Multiple Alignment	clustal Omega	Fast, accurate, scalable multiple sequence alignment for proteins
	clustalw	Multiple Sequence Alignment

Analysis Type	Application name	Description
clustalw2	Multiple Sequence Alignment	2.1
FAMSA	Progressive algorithm for large-scale multiple sequence alignments.	2.4.1
Gblocks	Gblocks eliminates poorly aligned positions and divergent regions of an alignment of DNA or protein sequences	0.91b
MAFFT	MAFFT is a multiple sequence alignment program	7.526
muscle	Multiple Sequence Alignment faster and more accurate than clustalw	5.1
t_coffee	Multiple sequence alignment package	12.00.7
Genome Aligner	minimap2	A versatile sequence alignment program that aligns DNA or mRNA se
	miniprot	Miniprot aligns a protein sequence against a genome with affine gap
Datbase search	dbget	DBGET is an integrated database retrieval system for major biologica
Sequence Assembler	CAP3	Multiple Sequence Alignment
	consed	Assembly Editor
	Phrap	Phrap is a program for assembling shotgun DNA sequence data
	Phred	The phred software reads DNA sequencing trace files, calls bases, an
	TGICL	Multiple Sequence Alignment (for huge data set)
Gene prediction	Augustus	AUGUSTUS is a program that predicts genes in eukaryotic genomic s
	Genemark	A family of gene prediction programs
	genscan	Gene prediction
	glimmer glimmerhmm	Glimmer is a system for finding genes in microbial DNA
	TSEBRA	TSEBRA is a combiner tool that selects transcripts from gene predicti
	Metaeuk	MetaEuk - sensitive, high-throughput gene discovery and annotation
	Prodigal	Prodigal: Fast, reliable protein-coding gene prediction for prokaryoti
Motif search	HMMER	Biosequence analysis using profile HMM

Analysis Type	Application name	Description
Interproscan	A tool that combines different protein signature recognition methods into one resource	5.75-106.0
meme	Multiple Em for Motif Elicitation	5.4.1
Functional annotation	eggNOG-Mapper	A tool for fast functional annotation of novel sequences.
phylogenetic tree analysis	mrbayes	MrBayes is a program for Bayesian inference and model choice acros
	njplot	NJplot is a tree drawing program
	paup	Tools for inferring and interpreting phylogenetic trees
	Phylip	A package of programs for inferring phylogenies (evolutionary trees)
	PhyML	PhyML is a phylogeny software based on the maximum-likelihood pri
	RAxML (raxmlHPC)	RAxML - Randomized Axelerated Maximum Likelihood
	tree-puzzle	Program to reconstruct phylogenetic trees from molecular sequence
Single cell analysis	CellRanger	A set of analysis pipelines that perform sample demultiplexing, barco analysis from single cell data.
	MARVEL	MARVEL is an R package developed for alternative splicing analysis a
Repeat Masking	RepeatMasker	RepeatMasker is a program that screens DNA sequences for interspe
	RepeatModeler	RepeatModeler is a de novo transposable element (TE) family identifi
tRNA search	tRNAscan-SE	tRNAscan-SE pioneers the large-scale use of covariance models to ar
Other	TransDecoder	TransDecoder identifies candidate coding regions within transcript s
	R	

? apptainer ????????????.sif ??????????????

- 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	O

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		ht
BRAKER	BRAKER is a program that predicts genes in eukaryotic genomic sequences	3.0.2	BRAKER/3.0.2/braker3.sif		ht
BUSCO	Assessing genome assembly and annotation completeness with single-copy orthologs	5.8.0	BUSCO/5.8.0/busco580.sif	Required Options : --offline	ht
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		ht
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		ht

EpiTyping	EpiTyping is a tool for detecting imprinting and X-chromosome inactivation status from RNA-seq	1	EpiTyping/epityping.sif		ht
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	ht
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		ht
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		ht
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		ht