Programas de bioinformatica instalados

1.0 Edition

Programas de bioinformatica instalados :

1.0 Edition
Published avril 25 2024

Table of Contents

List of Tables

Prologo

Ese documento presenta los programas de bioinformatica que se instalaron en el cluster, en la idea de proponer una ayuda a los que quisieran usarlos.

Chapter 1. Lista de los programas instalados

La lista siguiente muestra los programas que se instalaron ademas de los por default en el cluster. Checan si el programa que quieren utilizar contiene la columna module llena, o vacia.

Si contiene indicaciÃ³n de un module (por ejemplo el programa Samtools), vean como usar los modulos en la pagina siguiente.

Si no contiene informacion de module (por ejemplo ....), vean la pagina siiguiente "Uso de script file".

Los que necesiten utilizar **Python** Python o **R**, revisen las paginas siguientes relativas a estos dos programas, para conocer los paquetes instalados en ellos.


Revise las paginas siguientes para conocer como usar los programas aqui listados.

Table 1-1. Lista programas

Programa	Version	Tema	UbicaciÃ³n	
Abricate	0.9.8	Mass screening of contigs for antimicrobial resistance or virulence genes.	/share/apps/abricate-0.9.8	
Abyss (compilado con MPI)	2.1.5	Assemble large genomes using short reads.	/share/apps/Abyss-2.1.5-mpi	
Afterqc	0.9.7	Automatic Filtering, Trimming, Error Removing and Quality Control for fastq data.	/share/apps/AfterQC-0.9.7/	
Albacore	2.3.5	The standard basecaller package for NanoPore.	via conda..	pr
Allpaths-LG	52488	ALLPATHS-LG is a short-read assembler.	/share/apps/Allpaths-LG-52488/	
Amos	3.1.0-rc1	AMOS is a collection of tools and class interfaces for the assembly of DNA reads.	/share/apps/Amos-3.1.0-rc1	

Programa	Version	Tema	UbicaciÃ³n
Antismash	4.1.0	Search a genome sequence for secondary metabolite biosynthesis gene clusters	Modulo de Bioconda
"	7.0.1	" " "	Modulo de Bioconda
"			
"			
Anvio	7.1	an advanced analysis and visualization platform for omics data	Modulo de Conda
Anvio	8.0	an advanced analysis and visualization platform for omics data	Modulo de Conda
Aragorn	2.41	a program to detect tRNA genes and tmRNA genes in nucleotide sequence.	/share/apps/Aragorn-2.41
aria2	1.36.0	a lightweight multi-protocol & multi-source command-line download utility.	/share/apps/aria2-1.36.0
Art	2.5.8	ART is a set of simulation tools to generate synthetic next-generation sequencing reads.	/share/apps/Art-2.5.8/
Arumcrossversion	sin	a pipeline for analyzing and characterizing transcriptional readthrough as described in Roth et al.	paquage python 3
Atlas	2.1.4	Three commands to start analysing your metagenome data.	Module Python3
Atropos	1.1.31	a tool for specific, sensitive, and speedy trimming of NGS reads.	Module Python3

Programa	Version	Tema	UbicaciÃ³n
Autodock	4.2.6	AutoDock is a suite of automated docking tools.	/share/apps/autodock-4.2.6/
Autodock_vina	1.1.2	significantly improves the average accuracy of the binding mode predictions compared to AutoDock 4..	/share/apps/autodock-vina-1.1.2/
Bakta	1.9.2	.rapid & standardized annotation of bacterial genomes, MAGs & plasmids.	via conda
Bamtools	2.5.1	BamTools provides both a programmer's API and an end-user's toolkit for handling BAM files.	/share/apps/Bamtools-2.5.1/
BaseSpace-Cli	1.5.4	You can work with your BaseSpace Sequence Hub data using the command line interface (CLI).	/share/apps/BaseSpace-Cli-1.5.4
BayeScan	2.1	Detecting natural selection from population-based genetic data.	/share/apps/BayeScan-2.1/
BayesTraits	3.0.5	a computer package for performing analyses of trait evolution among groups of species for which a phylogeny or sample of phylogenies is available.	/share/apps/BayesTraits-3.0.5/
Basemount	0.15.2	BaseMount allows you to explore through runs, projects, samples, app results and analyses and interact directly with the associated files exactly as you would with any other file system.	Instalado via rpm
BBMap	38.26	BBMap is a splice-aware global aligner for DNA and RNA sequencing reads.	/share/apps/BBMap-38.26/
"	38.84	"	/share/apps/BBMap-38.84/
Beast2	2.4.5	BEAST 2 is a cross-platform program for Bayesian MCMC analysis of molecular sequences.	/share/apps/Beats-2.4.5/

Programa	Version	Tema	UbicaciÃ³n
Beast2	2.6.7	BEAST 2 is a cross-platform program for Bayesian MCMC analysis of molecular sequences.	/share/apps/Beats-2.6.7/
BEDops	2.4.40	The fast, highly scalable and easily-parallelizable genome analysis toolkit	/share/apps/Bedops-2.4.40/
Bcftools	1.9	Reading/writing BCF2/VCF/gVCF files and calling/filtering/summarising SNP and short indel sequence variants.	/share/apps/bcftools-1.9/
Bcftools	1.10.2	Reading/writing BCF2/VCF/gVCF files and calling/filtering/summarising SNP and short indel sequence variants.	/share/apps/bcftools-1.10.2/
Bcl2fastq	2.20	To demultiplex data and convert BCL files to FASTQ file formats.	/share/apps/bcl2fastq-2.20/
Bedtools2	2.26.0	The BEDTools utilities allow one to address common genomics tasks such as finding feature overlaps and computing coverage.	/share/apps/Bedtools2-2.26.0
Bedtools2	2.27.1	The BEDTools utilities allow one to address common genomics tasks such as finding feature overlaps and computing coverage.	/share/apps/Bedtools2-2.27.1
Bellerophon_version	sin version	a pipeline created to remove falsely assembled chimeric transcripts in de novo transcriptome assemblies.	Package Conda
bin3C	0.1.1	bin3C : Extract metagenome-assembled genomes (MAGs) from metagenomic data using Hi-C.	Paquete Python 2
BinSanity	0.2.7	BinSanity contains a suite a scripts designed to cluster contigs generated from a metagenomic assembly into putative genomes.	Paquete Python 2
Bismark	0.23.1	Bismark is a program to map bisulfite treated sequencing reads to a genome of interest and perform methylation calls in a single step.	/share/apps/Bismark-0.23.1/

Programa	Version	Tema	UbicaciÃ³n
Blast	2.7.1	The NCBI Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences.	/share/apps/Blast-2.7.1/
"	2.9.0	" " "	/share/apps/Blast-2.9.0/
"	2.11.0	" " "	/share/apps/Blast-2.11.0/
"	2.12.0	" " "	/share/apps/Blast-2.12.0/
Blat	35	BLAT (BLAST-like alignment tool) is a pairwise sequence alignment algorithm.	/share/apps/Blat-35/
Bonito	0.3.1	A PyTorch Basecaller for Oxford Nanopore Reads.	Paquete python3
Bowtie1	1.2.3	An ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.	/share/apps/Bowtie1-1.2.3/
Bowtie2	2.3.4.3	An ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.	/share/apps/Bowtie2-2.3.4.3/
"	2.5.0	" " "	/share/apps/Bowtie2-2.5.0/
"			
"			
Bracken	2.5	a highly accurate statistical method that computes the abundance of species in DNA sequences from a metagenomics sample.	/share/apps/Bracken-2.5/
Bracken2	2.1.6	an ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences.	Package Conda

Programa	Versión	Tema	Ubicación
BRB-seq	1.6	BRB-seq tools suite was created in the early days of multiplexed libraries, when there was not many other alternatives to analyze BRB-seq libraries..	/share/apps/BRB-seq-1.6/
Busco	4.0.5	BUSCO metric is complementary to technical metrics like N50	Package Conda
"	5.3.2	" " "	" " "
"			
"			
"	5.4.7	" " "	" " "
"			
"			
"			
Bwa	0.7.17	An efficient program that aligns relatively short nucleotide sequences against a long reference sequence such as the human genome.	/share/apps/Bwa-0.7.17
Canu	2.2	A single molecule sequence assembler for genomes large and small.	/share/apps/Canu-2.2
Casper	0.8.2	(Context-Aware Scheme for Paired-End Read) is state-of-the art merging tool in terms of accuracy and robustness.	/share/apps/Casper-0.8.2
CAT	4.6	tool for taxonomic classification of contigs and metagenome-assembled genomes (MAGs).	/share/apps/CAT-4.6
ccfind	1.4.5	a general tool to detect circular complete genomes with clues of terminal redundancy.	/share/apps/ccfind-1.4.5
Cd-hit	4.6.8	A very widely used program for clustering and comparing protein or nucleotide sequences.	/share/apps/Cdhit-4.6.8

Programa	Version	Tema	UbicaciÃ³n
Cd-hit	4.8.1	A very widely used program for clustering and comparing protein or nucleotide sequences.	/share/apps/Cdhit-4.8.1
cDNA_Cupcake	28.0	a miscellaneous collection of Python and R scripts used for analyzing sequencing data.	Paquage Conda
Cecret	3.7.20230620	Cecret was originally developed by @erinyoung at the Utah Public Health Laborotory for SARS-COV-2.	/share/apps/Cecret-3.7.20230620/
Checkm2	1.0.1	Rapid assessment of genome bin quality using machine learning..	Ambiance conda
Circlator	1.5.5	A tool to circularize genome assemblies.	Paquete Python 3
Clonalframeml	1.12	a software package that performs efficient inference of recombination in bacterial genomes.	Paquete Python 3
Comparem	0.1.1	a software toolkit which supports performing large-scale comparative genomic analyses.	via conda
Cufflinks	2.2.1	Transcriptome assembly and differential expression analysis for RNA-Seq.	/share/apps/Cufflinks-2.2.1
cutadapt	1.18	Cutadapt finds and removes adapter sequences, primers, poly-A tails and other types of unwanted sequence from your high-throughput sequencing reads.	Paquete Python 3
Deepbinner	0.2.0	Deepbinner is a tool for demultiplexing barcoded Oxford Nanopore sequencing reads.	Paquete Python 3
Deeploc (con licencia..)	2	.Prediction of eukaryotic protein subcellular localization using deep learning	Paquete conda

Programa	Versión	Tema	Ubicación
DeepMod	0.1.8	A deep-learning method for detecting DNA methylation state from Oxford Nanopore sequencing reads.	Paquete conda
DeepSimulator	1.0	The first deep learning based Nanopore simulator which can simulate the process of Nanopore sequencing.	/share/apps/DeepSimulator/
DeepTools	3.1.2	Tools for exploring deep sequencing data.	Paquete Python
DeepNano-blitz	0.1	a very fast basecaseller which can basecall reads as fast as they come from MinION on ordinary laptop.	Paquete conda
DeepVirFinder	-	predicts viral sequences using deep learning method..	Paquete Python 3
Diamond	0.9.22	A BLAST-compatible local aligner for mapping protein and translated DNA query sequences against a protein reference database (BLASTP and BLASTX alignment mode).	/share/apps/Diamond-0.9.22/
"	2.0.4	" "	/share/apps/Diamond-2.0.4
"	2.1.8	" "	/share/apps/Diamond-2.1.8
Discovar	52488	Discovar is a Genome variant discovery through assembly.	/share/apps/Discovar-52488/

Programa	Version	Tema	UbicaciÃ³n
Discovar de novo	52488	Discovar de novo is a new genome assembler.	/share/apps/Discovardenovo-52488/
DRAM	1.1.1	A tool for annotating metagenomic assembled genomes and VirSorter identified viral contigs.	via conda
DWGSIM	0.1.13	Whole Genome Simulator for Next-Generation Sequencing.	/share/apps/dwgsim-0.1.13
" " "	0.1.14	" " "	/share/apps/dwgsim-0.1.14
Edirect (NCBI Entrez Direct)	11.3	Entrez Direct (EDirect) provides access to the NCBI's suite of interconnected databases (publication, sequence, structure, gene, variation, expression, etc.) from a UNIX terminal window.	/share/apps/edirect-11.3
eggNOG mapper	2.1.2	a tool for fast functional annotation of novel sequences.	Module Python3
EIGENSTRAT	7.2.1	combines functionality from our population genetics methods and our EIGENSTRAT stratification correction method.	/share/apps/EIG-7.2.1/
Emboss	6.6.0	The European Molecular Biology Open Software Suite.	/share/apps/Emboss-6.6.0
Erne	2.1.1	The Extended Randomized Numerical alignEr using BWT.	/share/apps/Erne-2

Programa	Versión	Tema	UbicaciÃ³n
Ete3	3.1.1	A Python Environment for (phylogenetic) Tree Exploration.	Paquete Python 3
Exomiser	11.0.0	The Exomiser is a Java program that finds potential disease-causing variants from whole-exome or whole-genome sequencing data.	/share/apps/exomiser-cli-11.0.0
Exonerate	2.2.0	Exonerate is a generic tool for pairwise sequence comparison.	/share/apps/exonerate-2.2.0
"	2.4.0	" " "	/share/apps/exonerate-2.4.0
eXpress	1.5.1	eXpress is a streaming tool for quantifying the abundances of a set of target sequences from sampled subsequences.	/share/apps/express-1.5.1
FAPROTAX	1.2.4	Functional Annotation of Prokaryotic Taxa.	/share/apps/FAPROTAX-1.2.4
Fasta	36	find regions of local or global similarity between Protein or DNA sequences.	/share/apps/Fasta-36
FastANI	1.1	for fast alignment-free computation of whole-genome Average Nucleotide Identity (ANI).	/share/apps/FastANI-1.1/
"	1.33	" " "	/share/apps/FastANI-1.33/
"	1.34	" " "	/share/apps/FastANI-1.34/

Programa	Version	Tema	UbicaciÃ³n
Fastp	0.19.4	An ultra-fast all-in-one FASTQ preprocessor (QC/adapters/trimming/filtering/splitting...).	/share/apps/Fastp-0.19.4/
"	0.20.0	"	/share/apps/Fastp-0.20.0/
FastQC	0.11.7	A quality control tool for high throughput sequence data.	/share/apps/FastQC-0.11.7/
"	0.12.1	" " "	/share/apps/FastQC-0.12.1/
FastTree	2.1.10	FastTree infers approximately-maximum-likelihood phylogenetic trees from alignments of nucleotide or protein sequences.	/share/apps/FastTree-2.1.10/
"	2.1.11	" " "	/share/apps/FastTree-2.1.11/
FastViromeExplorer	..master..	Indentify the viruses/phages and their abundance in the viral metagenomics data.	/share/apps/FastViromeExplorer/
Fastx-Toolkit	0.13.0	a collection of command line tools for Short-Reads FASTA/FASTQ files preprocessing.	/share/apps/Fastx-Toolkit-0.13.0
Filtlong	0.2.0	Filtlong is a tool for filtering long reads by quality.	/share/apps/Filtlong-0.2.0/
FindFungi	0.23.3	A pipeline for the identification of fungi in public metagenomics datasets.	/share/apps/FindFungi-0.23.3/

Programa	Version	Tema	UbicaciÃ³n
Flappie	1.1.0	Flip-flop basecaller for Oxford Nanopore reads.	/share/apps/Flappie-1.1.0
Flye	2.6	De novo assembler for single molecule sequencing reads using repeat graphs.	/share/apps/Flye-2.6
" " "	2.9	" " "	/share/apps/Flye-2.9
fqtrim	0.9.7	filtering and trimming next generation sequencing reads.	/share/apps/fqtrim-0.9.7
Flash	1.2.11	FLASH (Fast Length Adjustment of SHort reads) is a very fast and accurate software tool to merge paired-end reads from next-generation sequencing experiments.	/share/apps/Flash-1.2.11
Focus	1.8.0	FOCUS: An Agile Profiler for Metagenomic Data	paquete Python 3
FragGeneScan	1.31	an application for finding (fragmented) genes in short reads.	/share/apps/FragGeneScan/
Freebayes	1.2.0	Bayesian haplotype-based polymorphism discovery.	/share/apps/Freebayes-1.2.0/
" " "	1.3.6	" " "	/share/apps/Freebayes-1.3.6/
FUSTr	sin version	a pipeline that clusters coding sequences from transcriptomes into protein families.	/share/apps/FUSTr

Programa	Version	Tema	UbicaciÃ³n
GapFiller	2.1.1	A seed-and-extend local assembler to fill the gap within paired reads.	/share/apps/Gapfiller-2.1.1/
GATK	4.1.4.1	Variant Discovery in High-Throughput Sequencing Data.	/share/apps/gatk-4.1.4.1
GeneMark-ES	1	A family of gene prediction programs (Bajo licencia!)	/share/apps/GeneMark-ES
UCSC Genome Browser Utilities	s/n	Algunas herramientas : bigWigMerge faToTwoBit twoBitInfo twoBitToFa .	/share/apps/UCSC_Utilities
GenomeMapper	0.4.4	GenomeMapper is a short read mapping tool designed for accurate read alignments.	/share/apps/GenomeMapper-0.4.4
get_homologues	3.2	a versatile software package for pan-genome analysis.	/share/apps/get_homologues-3.2
get_phylomarkers	1.2.9	A pipeline to select optimal markers for microbial phylogenomics, population genetics and genomic taxonomy.	/share/apps/get_phylomarkers-1.2.9
get_phylomarkers	2.2.9.1	A pipeline to select optimal markers for microbial phylogenomics, population genetics and genomic taxonomy.	/share/apps/get_phylomarkers-2.2.9.1
Gffcompare	0.10.6	Classify, merge, tracking and annotation of GFF files by comparing to a reference annotation GFF.	/share/apps/Gffcompare-0.10.6/
Glimmer	3.02	A system for finding genes in microbial DNA, especially the genomes of bacteria, archaea, and viruses.	/share/apps/Glimmer-3.02/

Programa	Versión	Tema	Ubicación
GlimmerHMM	3.0.4	A new gene finder based on a Generalized Hidden Markov Model (GHMM).	/share/apps/GlimmerHMM-3.0.4/
Grinder	0.5.4	Grinder is a versatile open-source bioinformatic tool to create simulated omic shotgun and amplicon sequence libraries for all main sequencing platforms.	/share/apps/Grinder-0.5.4/
GraftM	0.13.1	GraftM : a tool for finding genes of interest in metagenomes, metatranscriptomes, and whole genomes.	paquete Python 3
GraftM	0.14.0	GraftM : a tool for finding genes of interest in metagenomes, metatranscriptomes, and whole genomes.	/share/apps/graftM-0.14.0 y ambiente conda...
Graphlan	1.1.3	GraPhlAn is a software tool for producing high-quality circular representations of taxonomic and phylogenetic trees.	/share/apps/graphlan-1.1.3
Gromacs (version MPI)	5.0.7	Molecular dynamics software package mainly designed for simulations of proteins, lipids and nucleic acids.	/share/apps/Gromacs-5.0.7/
" "	2020.3	" "	/share/apps/Gromacs-2020.3/
GTDB-Tk	2.1.0	a software toolkit for assigning objective taxonomic classifications to bacterial and archaeal genomes.	Module Python 3.9
GToTree	1.4.2	User-friendly workflow for phylogenomics intended to give more researchers the capability to create phylogenomic trees..	via conda..
Guppy	6.5.7	" "	/share/apps/ont-guppy-cpu-6.5.7/

Programa	Version	Tema	UbicaciÃ³n
HiCExplorer	3.7.2	Set of programs to process, normalize, analyze and visualize Hi-C and HiChIP data. Atencion: La version instalada es via un contenedor Singulariy, verificar la documentacion del mudule!!!	/share/apps/HicExplorer-3.7.2/
HiC-PRO	3.0.0	An optimized and flexible pipeline for Hi-C data processing.	/share/apps/HiC-Pro_3.0.0
Hisat2	2.1.0	Graph-based alignment of next generation sequencing reads to a population of genomes.	/share/apps/Hisat2-2.1.0
Hmmer v2	2.3.2	Profile hidden Markov models for biological sequence analysis.	/share/apps/Hmmer-2.3.2/
Hmmer v3	3.2.1	Profile hidden Markov models for biological sequence analysis.	/share/apps/Hmmer-3.2.1/
Hmmer v3	3.3.2	Profile hidden Markov models for biological sequence analysis.	/share/apps/Hmmer-3.3.2/
Homer	4.10	Motif discovery and next generation sequencing analysis	/share/apps/Homer-4.10/
HumanN	3.0.0a4	next iteration of HUMAnN, the HMP Unified Metabolic Analysis Network.	Ambiante conda
IQ-Tree	1.6.12	Efficient software for phylogenomic inference.	/share/apps/iqtree-1.6.12/
ivar	1.0	A computational package that contains functions broadly useful for viral amplicon-based sequencing.	/share/apps/ivar-1.0/

Programa	Versiona	Tema	UbicaciÃ³n
"	1.3.1	" " "	/share/apps/ivar-1.3.1/
Idba	1.1.3	a iterative De Bruijn Graph De Novo Assembler for Short Reads Sequencing data with Highly Uneven Sequencing Depth.	/share/apps/Idba-1.1.3/
Idba-mt	1.0	is a post-processing software for IDBA-UD contigs for removing chimeria contigs and extending contig length using paired-end reads information.	/share/apps/Idba-MT-1.0
Infernal	1.1.3	for searching DNA sequence databases for RNA structure and sequence similarities.	/share/apps/infernal-1.1.3/
Interproscan	5.51-85.0	a database which integrates together predictive information about proteins' function from a number of partner resources, giving an overview of the families that a protein belongs to and the domains and sites it contains..	/share/apps/interproscan-5.51-85.0
"	5.55-88.0	" " "	/share/apps/interproscan-5.51-88.0
ITSxpress	2.0.0	Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis.	Package Conda
Jellyfish	2.2.10	A fast multi-threaded k-mer counter	/share/apps/jellyfish-2.2.10/
jModelTest	2.1.10	a tool to carry out statistical selection of best-fit models of nucleotide substitution.	/share/apps/jModelTest-2-1.10/
Kaiju	1.7.3	a program for the taxonomic classification of high-throughput sequencing reads.	/share/apps/kaiju-1.7.3/

Programa	Version	Tema	UbicaciÃ³n
Kaiju	1.8.0	" "	/share/apps/kaiju-1.8.0/
Kallisto	0.44.0	A program for quantifying abundances of transcripts from RNA-Seq data.	/share/apps/Kallisto-0.44.0/
Kraken 2	2.0.9	a taxonomic sequence classifier.	/share/apps/Kraken-2.0.9/
" " "	2.1.2	" " "	/share/apps/Kraken2-2.1.2/
Kraken-biom	1.0.1	Create BIOM-format tables from Kraken output.	Module python 3
LongQC	1.2.0c	a tool for the data quality control of the PacBio and ONT long reads, and it has two functionalities: sample qc and platform qc.	/share/apps/LongQC-1.2.0c
Macs	3.0.0a6	Model-based Analysis for ChIP-Seq	Module python 3
Mafft	7.407	Multiple alignment program for amino acid or nucleotide sequences	/share/apps/Mafft-7.407
" " "	7.505	" " "	/share/apps/Mafft-7.505
Manorm	??	A robust model for quantitative comparison of ChIP-Seq data sets.	/share/apps/MAnorm

Programa	Versiones	Tema	UbicaciÃ³n
Mauve	2015-02-13	Multiple Alignment of Conserved Genomic Sequence with Rearrangements.	/share/apps/mauve-2015-02-13
MASH	2.3	Fast genome and metagenome distance estimation using MinHash.	/share/apps/Mash-2.3
Masurca	3.3.4	The MaSuRCA (Maryland Super Read Cabog Assembler) assembler combines the benefits of deBruijn graph and Overlap-Layout-Consensus assembly approaches.	/share/apps/Masurca-3.3.4
Masurca	3.4.2	" "	/share/apps/Masurca-3.4.2
MaxBin	2.2.7	MaxBin is software for binning assembled metagenomic sequences based on an Expectation-Maximization algorithm.	/share/apps/MaxBin-2.2.7
Medaka	0.11.5	Sequence correction provided by ONT Research.	Module Python3
Megahit	1.1.3	An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.	/share/apps/Megahit-1.1.1
Meme	4.11.3	Perform motif discovery on DNA, RNA, protein or custom alphabet datasets.	/share/apps/Meme-4.11.3
Metabat	236d20e76303	A robust statistical framework for reconstructing genomes from metagenomic data.	/share/apps/MetaBat-236d20e76303/
MetaCluster	5.1	a new software for binning short pair-end reads.	/share/apps/MetaCluster-5.1/
Metaerg	1.2.3	a stand-alone and fully automated metagenomic and metaproteomic data annotation pipeline.	/share/apps/Metaerg-1.2.3/

Programa	Version	Tema	UbicaciÃ³n
Metaphlan v3	3.0.2	Metagenomic Phylogenetic Analysis.	Ambiente Conda
Metaphlan v4	4.0.2	Metagenomic Phylogenetic Analysis.	Ambiente Conda
MetaBAT	1.1	an Metagenomic Scanning and Annotation tool, with an emphasis on metabolic genes.	Ambiente Conda ligado con prokka
MetaVelvet-SL	1.0	An extension of Velvet assembler to de novo metagenomic assembler utilizing supervised learning.	/share/apps/MetaVelvet-SL-1.0/
MetaWRAP	1.3	aims to be an easy-to-use metagenomic wrapper suite that accomplishes the core tasks of metagenomic analysis from start to finish.	Ambiente Conda
MetReTrim	sin version	a pipeline, written in python, for trimming heterogeneity 'N' spacers from the pre-processed reads given the primer sequences.	/share/apps/MetReTrim/
Miniasm	0.3-r179	Ultrafast de novo assembly for long noisy reads (though having no consensus step).	/share/apps/Miniasm-0.3-r179
Minimap2	2.14-r883	A versatile pairwise aligner for genomic and spliced nucleotide sequences.	/share/apps/Minimap2-2.14-r883
Minimap2	2.17	"	/share/apps/Minimap2-2.17
MinPath	1.6	a parsimony approach for biological pathway reconstructions using protein family predictions.	/share/apps/MinPath-1.6

Programa	Versión	Tema	Ubicación
Mira	4.0.2	Sequence assembler and sequence mapping for whole genome shotgun and EST / RNASeq sequencing data.	/share/apps/mira-4.0.2
miRanda	3.3a	an algorithm for finding genomic targets for microRNAs.	Module conda
Mirdeep2	0.1.3	Discovering known and novel miRNAs from deep sequencing data.	/share/apps/mirdeep2-0.1.3
MMseqs2	13-45111	a software suite to search and cluster huge sequence sets.	/share/apps/MMseqs2-13-45111
" " "	14-7e284	" " "	/share/apps/MMseqs2-14-7e284
Mob-suite	3.1.8	via conda.	/share/apps/Modeller-9.21
Modeller	9.21	For homology or comparative modeling of protein three-dimensional structures (1,2).	/share/apps/Modeller-9.21
Mosaik	2.2.30	reference-guided aligner for next-generation sequencing technologies.	/share/apps/Mosaik-2.2.30
MrBayes	3.2.6	MrBayes is a program for Bayesian inference and model choice across a large space of phylogenetic and evolutionary models.	/share/apps/MrBayes-3.2.6
Mummer	3.23	Ultra-fast alignment of large-scale DNA and protein sequences.	/share/apps/Mummer-3.23
Muscle	3.8.1551	Multiple sequence alignment. Faster and more accurate than CLUSTALW.	/share/apps/Muscle-3.8.1551

Programa	Version	Tema	UbicaciÃ³n
Mycoscan	2.0	Characterisation of fungal fraction in metagenomic surveys, with high end-user flexibility in terms of database creation and customisation.	/share/apps/MScan-2.0
Nanoplot	1.19.0	Plotting tool for long read sequencing data and alignments.	Paquete python 3
Nanopolish	1.10.2	Signal-level algorithms for MinION data	/share/apps/Nanopolish-0.10.2
Nextalign	0.1.6	Nextalign is a sequence reference alignment tool which uses the same alignment algorithm as Nextclade	/share/apps/nextalign-0.1.6
Nextalign	1.11.0	Nextalign is a sequence reference alignment tool which uses the same alignment algorithm as Nextclade	/share/apps/nextalign-1.11.0
Nextclade	0.13.0	Nextclade : Clade assignment, mutation calling, and sequence quality checks	En Node.js
Nextclade	1.11.0	Nextclade : Clade assignment, mutation calling, and sequence quality checks	/share/apps/nextclade-1.11.0
NGSutils	0.5.9	NGSUtils : Tools for next-generation sequencing analysis.	Ambiente Conda
Nonpareil	3.4.1	Nonpareil uses the redundancy of the reads in metagenomic datasets to estimate the average coverage and predict the amount of sequences that will be required to achieve nearly complete coverage.	Ambiente Conca
Nseg	?	Low complexity sequence identification.	/share/apps/nseg
Octave	3.8.2	high-level language, primarily intended for numerical computations.	Instalado via RPM

Programa	Version	Tema	UbicaciÃ³n
OpenRDP	v0.1a0	an open-source re-implementation of the RDP recombination detection program.	Programa Python
OrthoFinder	2.5.5	a fast, accurate and comprehensive platform for comparative genomics.	/share/apps/OrthoFinder-2.5.5
Pandaseq	2.11	a program to align Illumina reads, optionally with PCR primers embedded in the sequence, and reconstruct an overlapping sequence.	/share/apps/pandaseq-2.11
Parallel-META	2.4.1	Parallel-META is a a comprehensive and full-automatic computational toolkit for rapid data mining among metagenomic datasets.	/share/apps/parallel-meta-2.4.1/
" " "	3.5.3	" " "	/share/apps/parallel-meta-3.5.3/
" " "	3.7.0	" " "	/share/apps/parallel-meta-3.5.3/
PATT	0.0.1	Proteome Annotation Transfer Tool (pipeline de la UUSMB).	/share/apps/PATT
pb-assembly	0.0.8	.	dentro de miniconda
Picard	2.18.14	a set of tools (in Java) for working with next generation sequencing data in the BAM format..	/share/apps/picard-tools-2.18.14
Picrust2	2.3.0_b	software for predicting functional abundances based only on marker gene sequences.	dentro de miniconda

Programa	Version	Tema	UbicaciÃ³n
Picrust2 (plugin qiime2)	2.3.0_b	software for predicting functional abundances based only on marker gene sequences.	plugin qiime2
Pigz	2.4	A parallel implementation of gzip for modern multi-processor, multi-core machines.	/share/apps/External/pigz-2.4
Pilon	1.23	Automatically improve draft assemblies / Find variation among strains, including large event detection	/share/apps/Pilon-1.23
" " "	1.24	" " "	/share/apps/Pilon-1.24
Platanus	1.2.4	PLATform for Assembling NUcleotide Sequences	/share/apps/Platanus-1.2.4
Plink	1.9	free, open-source whole genome association analysis toolset, designed to perform a range of basic, large-scale analyses in a computationally efficient manner.	/share/apps/plink-1.9
Pomoxis	0.2.3	Pomoxis comprises a set of basic bioinformatic tools tailored to nanopore sequencing.	dentro de miniconda
Porechop	0.2.4	Porechop is a tool for finding and removing adapters from Oxford Nanopore reads.	/share/apps/Porechop-0.2.4/
pplacer	1.1.alpha19	Phylogenetic placement and downstream analysis.	/share/apps/pplacer-1.1.alpha19
Pratt	2.1	A tool for finding flexible patterns in unaligned protein sequences.	/share/apps/Pratt-2.1

Programa	Versiona	Tema	UbicaciÃ³n	
Preseq	2.0.1	The preseq package is aimed at predicting and estimating the complexity of a genomic sequencing library	/share/apps/preseq-2.0.1/	
Prodigal	2.6.3	Fast, reliable protein-coding gene prediction for prokaryotic genomes.	/share/apps/Prodigal-2.6.3	
Proteinortho	6.0.33	a tool to detect orthologous genes within different species.	/share/apps/proteinortho-6.0.33	
Prottest3	3.4.2	a bioinformatic tool for the selection of best-fit models of aminoacid replacement for the data at hand.	/share/apps/prottest3-3.4.2	
Prokka	1.14.6	rapid prokaryotic genome annotation.	/share/apps/prokka-1.14.6	
pyEGA3	5.0.2	The pyEGA3 download client is a python-based tool for viewing and downloading files from authorized EGA datasets.	paquete python	
Qiime2	2019.4	a next-generation microbiome bioinformatics platform.	via conda..	pr
Qiime2	2019.10	a next-generation microbiome bioinformatics platform.	via conda..	pr
Qiime2	2022.8	a next-generation microbiome bioinformatics platform.	via conda..	pr
Quast (MetaQuast)	5.0.2	evaluates genome assemblies by computing various metrics.	Paquete Python 3	
R Project	3.5.1	R is a free software environment for statistical computing and graphics.	/share/apps/R-3.5.1/	

Programa	Version	Tema	UbicaciÃ³n
R Project	4.0.3	R is a free software environment for statistical computing and graphics.	/share/apps/R-4.0.3/
R Project	4.1.2	R is a free software environment for statistical computing and graphics.	/share/apps/R-4.1.2/
Racon	1.3.1	Consensus module for raw de novo DNA assembly of long uncorrected reads.	/share/apps/Racon-1.3.1
Randfold	2.0.1	Minimum free energy of folding randomization test software.	/share/apps/Randfold-2.0.1
Ratatosk	0.2	Hybrid error correction of long reads using colored de Bruijn graphs	/share/apps/Ratatosk-0.2
RAxML	8.2.12	Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Utiliza 2 threads por defaults!	/share/apps/RAxML-8.2.12
Rcorrector	1.0.4	Efficient and accurate error correction for Illumina RNA-seq reads.	/share/apps/Rcorrector-1.0.4
Recon	1.08	a package for automated de novo identification of repeat families from genomic sequences.	/share/apps/Recon-1.08
RepeatMasker	4.0.8	program that screens DNA sequences for interspersed repeats and low complexity DNA sequences.	/share/apps/RepeatMasker-4.0.8
" " "	4.2.1	" " "	/share/apps/RepeatMasker-4.2.1

Programa	Versión	Tema	Ubicación
RepeatModeler	1.0.11	a de novo repeat family identification and modeling package.	/share/apps/RepeatModeler-1.0.11
"	2.0.3	" " "	/share/apps/RepeatModeler-2.0.3
"			
"			
RepeatSout	1.0.5	identify repeat family sequences from genomes where hand-curated repeat databases (a la RepBase update) are not available.	/share/apps/RepeatSout-1.0.5
RMBlastn	2.6.0	RMBlast is a RepeatMasker compatible version of the standard NCBI blastn program.	/share/apps/rmblast-2.6.0
RNAhybrid	2.1.2	RNAhybrid is a tool for finding the minimum free energy hybridization of a long and a short RNA.	/share/apps/RNAhybrid-2.1.2
Rnammer	1.2	RNAmmer predicts 5s/8s, 16s/18s, and 23s/28s ribosomal RNA in full genome sequences.	/share/apps/rnammer-1.2
Roary	3.13.0	The pan genome pipeline.	Ambiante conda
Rsem	1.3.1	RSEM is a software package for estimating gene and isoform expression levels from RNA-Seq data.	/share/apps/Rsem-1.3.1
Salmon	0.11.3	Salmon is a tool for wicked-fast transcript quantification from RNA-seq data.	/share/apps/Salmon-0.11.3/
Samtools	0.1.18	Flexible generic format for storing nucleotide sequence alignment	/share/apps/samtools-0.1.18/
Samtools	1.9	" " "	/share/apps/samtools-1.9/

Programa	Versión	Tema	UbicaciÃ³n
$samtools	1.10	" " "	/share/apps/samtools-1.10/
$SCRMshaw_HD	git version	Genome-wide CRM Prediction Program	/share/apps/SCRMshaw_HD.git/
$SEACR	1.4beta	SEACR is intended to call peaks and enriched regions from sparse CUT and RUN or chromatin profiling data in which background is dominated by "zeroes" (i.e. regions with no read coverage).	/share/apps/SEACR-1.4.b2/
seq_scrumbs	0.1.9	Little sequence file utilities meant to work within Unix pipelines	Paquete Python2
$seqkit	0.9.1	A cross-platform and ultrafast toolkit for FASTA/Q file manipulation in Golang	/share/apps/Seqkit-0.9.1
" "	0.13.2	" " "	/share/apps/Seqkit-0.13.2
" " "	2.5.1	" " "	/share/apps/Seqkit-2.5.1
$shasta	0.5.1	The goal of the Shasta long read assembler is to rapidly produce accurate assembled sequence using as input DNA reads generated by Oxford Nanopore flow cells.	/share/apps/shasta-0.5.1
$Shore	0.9.3	SHORE is a mapping and analysis pipeline for short DNA sequences.	/share/apps/Shore-0.9.3
$ShoreMap	3.6	ShoreMap : For fast and accurate identification of causal mutations in plants.	/share/apps/ShoreMap-3.6

Programa	Versiones	Tema	UbicaciÃ³n
`$signalp` (requiere licencia)	4.1	SignalP predicts the presence and location of signal peptide cleavage sites in amino acid sequences from different organisms.	/share/apps/signalp-4.1
"	5.0b	" " " " "	/share/apps/signalp-5.0b
"	6.0g	" " " " "	Ambiente conda
`$skewer`	0.2.2	implements the bit-masked k-difference matching algorithm dedicated to the task of adapter trimming and it is specially designed for processing next-generation sequencing (NGS) paired-end sequences.	/share/apps/skewer-0.2.2/
`$smalt`	0.7.6	A mapper for DNA sequencing reads.	/share/apps/Smalt-0.7.6
`$snap`	2.0.1	Scalable Nucleotide Alignment Program.	/share/apps/snap-2.0.1
`$snakemake`	5.3.0	aims to reduce the complexity of creating workflows by providing a fast and comfortable execution environment.	Paquete Python 3
"	5.5.0	"	"
"	5.10.0	"	"
`$snippy`	4.6.0	Snippy finds SNPs between a haploid reference genome and your NGS sequence reads	/share/apps/snippy-4.6.0/

Programa	Versiones	Tema	UbicaciÃ³n
`$ourmash`	3.0.1	Compute and compare MinHash signatures for DNA data sets.	Paquete Python3
`spacegraphcats`	1.2	Explore large, annoying graphs using hierarchies of dominating sets.	/share/apps/spacegraphcats
`SPAdes`	3.13.0	St. Petersburg genome assembler : is an assembly toolkit containing various assembly pipelines.	/share/apps/SPAdes-3.13.0
"	3.14.0	"	/share/apps/SPAdes-3.14.0
"	3.14.1	"	/share/apps/SPAdes-3.14.1
"	3.15.4	"	/share/apps/SPAdes-3.15.4
`$queezeMeta`	1.4.0	A complete pipeline for metagenomic analysis.	/share/apps/SqueezeMeta-1.4.0
" " "	1.5.2	A complete pipeline for metagenomic analysis.	/share/apps/SqueezeMeta-1.5.2
`sra-tools`	3.0.0	The SRA Toolkit and SDK from NCBI is a collection of tools and libraries for using data in the INSDC Sequence Read Archives.	/share/apps/sra-tools-3.0.0
`Stacks`	2.62	a software pipeline for building loci from short-read sequences, such as those generated on the Illumina platform.	/share/apps/stacks-2.62

Programa	Versiona	Tema	UbicaciÃ³n
	2.66	" " "	/share/apps/stacks-2.66
	2.7.10a	Spliced Transcripts Alignment to a Reference.	/share/apps/Star-2.7.10a
STAR			
	2.7.11a	" " "	/share/apps/Star-2.7.11a
	0.34	A tool for agile functional analysis of shotgun metagenomic data.	via conda
Super-Focus			
	5.2.60	Trait Analysis by aSSociation, Evolution and Linkage.	/share/apps/tassel-5.2.60
Tassel			
	2.0c	Prediction of transmembrane helices in proteins.	/share/apps/tmhmm-2.0c
Tmhmm			
	5.3.0	Find Coding Regions Within Transcripts.	/share/apps/TransDecoder-5.3.0/
TransDecoder			
	1.3	.an efficient de novo trascriptome assembler for RNA-Seq data.	/share/apps/Translig-1.3/
TransLig			
	4.09	.	/share/apps/trf-4.09/
Tandem Repeats Finder (TRF)			

Programa	Version	Tema	UbicaciÃ³n
TrimGalore	0.6.7	a wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data.	/share/apps/TrimGalore-0.6.7
Trimmomatic	0.38	Trimmomatic is a fast, multithreaded command line tool that can be used to trim and crop Illumina (FASTQ) data as well as to remove adapters.	/share/apps/Trimmomatic-0.38/
"	0.39	" " "	/share/apps/Trimmomatic-0.39/
Trinity	2.8.4	RNA-Seq de novo transcriptome assembly.	/share/apps/trinityrnaseq-2.8.4/
"	2.9.1	"	/share/apps/trinityrnaseq-2.9.1/
"	2.15.1	"	/share/apps/trinityrnaseq-2.15.1/
Trinotate	3.1.1	Transcriptome Functional Annotation and Analysis	/share/apps/Trinotate-3.1.1/
UMI-tools	1.1.4	Tools for dealing with Unique Molecular Identifiers.	Paquete Python 3
Unicycler	0.4.8	an assembly pipeline for bacterial genomes.	Paquete Python 3
VarSeq Tools	3.0.2	an integrated annotation and analysis package for next-generation sequencing data.	Paquete Python 3

Programa	Version	Tema	UbicaciÃ³n
Vcftools	0.1.16	a program package designed for working with VCF file.	/share/apps/Vcftools-0.1.16
vConTACT2	0.9.11	a tool to perform guilt-by-contig-association classification of viral genomic sequence data.	Paquete Python 3
Verkko	1.3.1	a hybrid genome assembly pipeline developed for telomere-to-telomere assembly of PacBio HiFi and Oxford Nanopore reads.	Paquete Conda
VIRAS	sin version	a Viral genome Evolution Network Analysis System.	/share/apps/Venas
Velvet	1.2.10	Sequence assembler for very short reads.	/share/apps/Velvet-1.2.10
Vibrant	1.2.1	Virus Identification By iteRative ANnoTation.	/share/apps/Vibrant-1.2.1/
ViennaRNA	2.4.13	The ViennaRNA Package consists of a C code library and several stand-alone programs for the prediction and comparison of RNA secondary structures.	/share/apps/ViennaRNA-2.4.13/
VirSorter	1.0.5	mining viral signal from microbial genomic data	En conda. Leer la documentacion adjunta del module (module help programs/virsorter-1.0.5) .
Vsearch	2.13.4	Versatile open-source tool for metagenomics.	/share/apps/Vsearch-2.13.4/
Wub	0.3.0	Tools and software library developed by the ONT Applications group.	Paquete python 3

Chapter 2. Uso de los programas

2.1. Module

La mayoria de los programas ya cuentan con la facilidad de usarlos via el comando **module**. Este comando permite configurar variables de ambientes para los programas, pero ademas permite desconfigurarlas de la misma manera...

En una session normal, podran ver los modules existente que estan usando:

```
$ module list
Currently Loaded Modulefiles:
1) rocks-openmpi
```

Se puede averiguar los modulos disponibles (lista no actualizada, es para ejemplo practico):

```
$ module avail

../..

------------------------------------------------------------ /share/apps/Modules ------------------
compilers/cmake-3.12.2  compilers/python-3.6.6  programs/bamtools-2.5.1 programs/freebayes      p
```

Para tener un poco mas de informacion sobre un programa con modulo, se puede utilizar el sub-comando help:

```
$  module help programs/R-3.5.1

----------- Module Specific Help for 'programs/R-3.5.1' -----------

 R is a free software environment for statistical computing and graphics.
```

La columna "Module" en la lista completa de los programas disponibles indica el nombre del modulo a utilizar.

!IMPORTANTE! Para usar un module en un script, deben de escribir esta linea antes de usarlo:

```
#$ .... opciones de SGE

source $HOME/.bashrc
module load modulename
```

El problema es debido a que no se genera una coneccion con ejecucion normal del Shell. Estamos revisando como corregir este problema.

2.2. Uso de script file

Los programas que no cuentan con un modulo propio, habra que definir por lo menos la ruta de acceso a ellos mismo. En ese sentido se genero archivos que permiten facilitar la definicion de esas rutas.

Dependiendo de la ubicacion de programa que quiserian usar, es recomendable incluir en su script de jobs, al inicio, las siguientes lineas:

```
source ~/.bashrc

source /share/apps/Profiles/share-profile.sh
```

2.3. Python 2 y 3

Se instalaron las 2 versiones de Python, debido a que todavia la verions 3 de python no es toitalmente aceptada por todos los programas.

Para usar **Python2**, cargan el modulo correspondiente: **compilers/python-2.7.15**. Una vez cargado podran detectar y ver la lista de los modulos disponibles:

```
$ module load compilers/python-2.7.15

$ pip2 list
Package        Version
-------------- -----------
appdirs        1.4.3
certifi        2018.8.24
chardet        3.0.4
../..
```

Para usar **Python3**, cargan el modulo correspondiente: **compilers/python-3.6.6**. Una vez cargado podran detectar y ver la lista de los modulos disponibles:

```
$ module load compilers/python-3.6.6

$ pip3 list
Package        Version
-------------- -----------
appdirs        1.4.3
certifi        2018.8.24
chardet        3.0.4
../..
```

2.4. Perl

Se instalo la version mÃ¡s reciente de **Perl**, la 5.28. Con esta version se instalaron los modules como Bio::Perl.

Para utilizar esta version, cargan el modulo de perl :

```
$ which perl
/usr/bin/perl
$ module load compilers/perl-5.28
$ which perl
/share/apps/External/Perl-5.28/bin/perl
```

Si requieren un modulo en especifico, pidenlo para que se integre en esta versiÃ³n de Perl.

2.5. R

Se instalo la version 3.5.1 de R en el cluster. Para utilizarl, se debe de cargar el modulo correspondiente:

```
$ module load programs/R-3.5.1

$  R

R version 3.5.1 (2018-07-02) -- "Feather Spray"
Copyright (C) 2018 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

Para ver los paquetes disponibles en R (la lista aqui no esta actualizada..)

```
>installed.packages()[,"Package"]
           BH           DBI    KernSmooth          MASS        Matrix
         "BH"         "DBI"  "KernSmooth"        "MASS"      "Matrix"
           R6          Rcpp         abind       acepack         akima
         "R6"        "Rcpp"       "abind"     "acepack"       "akima"
          ape    assertthat     backports          base     base64enc
        "ape"  "assertthat"   "backports"        "base"   "base64enc"
       bibtex         bindr      bindrcpp           bit         bit64
     "bibtex"       "bindr"    "bindrcpp"         "bit"       "bit64"
       bitops          blob          boot       caTools       carData
     "bitops"        "blob"        "boot"     "caTools"     "carData"
```

```
   cellranger     checkmate         chron         class           cli
 "cellranger"   "checkmate"       "chron"       "class"         "cli"
      cluster     codetools    colorspace      combinat      compiler
    "cluster"   "codetools"  "colorspace"    "combinat"    "compiler"
       crayon          curl    data.table      datasets     dichromat
     "crayon"        "curl"  "data.table"    "datasets"   "dichromat"
       digest          doMC    doParallel         dplyr         fansi
     "digest"        "doMC"  "doParallel"       "dplyr"       "fansi"
      foreach       foreign          glue     grDevices      graphics
    "foreach"     "foreign"        "glue"   "grDevices"    "graphics"
         grid     iterators       lattice      magrittr       methods
       "grid"   "iterators"     "lattice"    "magrittr"     "methods"
        mgcv          nlme          nnet      parallel        pillar
       "mgcv"        "nlme"        "nnet"    "parallel"      "pillar"
    pkgconfig         plogr   prettyunits         purrr       rematch
  "pkgconfig"       "plogr" "prettyunits"       "purrr"     "rematch"
       rlang         rpart            sp       spatial       splines
     "rlang"       "rpart"          "sp"     "spatial"     "splines"
       stats        stats4       stringi       stringr      survival
     "stats"      "stats4"     "stringi"     "stringr"    "survival"
       tcltk        tibble    tidyselect         tools          utf8
     "tcltk"      "tibble"  "tidyselect"       "tools"        "utf8"
       utils
     "utils"
```

2.6. Java (JRE y JDK)

Para poder correr programas escritos en **Java**, deberan de utilizar un modulo en especifico. Por ejemplo, para utilizar el java en la versiÃ³n 1.8 **1.8**, se utiliza el modulo correspondiente, como se lo indica a continuaciÃ³n:

```
$ module load compilers/jre1.8.0_181
$ which java
/share/apps/External/jre1.8.0_181/bin/java
```

Para conocer las versiones disponibles, utilzan el comando module de la manera siguiente:

```
$ module avail compilers/j

---------------------- /share/apps/Modules ------------------------------------------------------
compilers/jdk-11.0.5   compilers/jdk1.8.0_181 compilers/jre1.8.0_181
```

Normalmente, para los programas instalados y listados en la pagina del cluster, que necesiten java, como **FastQC**, su modulo incluye de manera automatica la puesta en marcha del modulo java correspondiente.

Para compilar programas escritos en Java, utilizan el modulo **compilers/jdk1.8.0_181**

```
$ module load compilers/jdk1.8.0_181
$ which javac
/share/apps/External/jdk1.8.0_181/bin/javac
```

2.7. Programas especiales

El programa **AfterQC** necesita Python. Se debe de utilizar con el modulo adecuado, como se muestra aqui:

```
$ module load programs/afterqc-0.9.7

$ after.py
specify current dir as input dir
no read files to run with, do you call the program correctly?
see -h for help
Time used: 0.00595998764038
```

Para las opciones revisen por favor la documentacion de AfterQC o directamente en linea[1].

El programa **Trinity** en su version mÃ¡s reciente necesita por default mÃ¡s memoria que los 7G de base en casa jobs. Por ello se recomienda utilizarlo con al menos 32 Gb con la opciÃ³n:

```
      -l h_vmem=32G
```

El programa **Albacore** necesita un ambiente especial para poder correr, dentro de Conda. Una vez cagado el modulo de miniconda, es necesario cargar el ambiene de albacore:

```
$ module load programs/albacore-2.3.1
$ source activate albacore
(albacore) $  --> Indica que estan en el ambiente de albacore, listo!
```

Para quitar el ambiente de trabajo de Ablacore y poder seguir trabajando (en el mismo job..), es necesario realizar las dos operaciones siguientes:

```
(albacore) $ source deactivate albacore
$ module unload programs/albacore-2.3.1
```

El procedimiento parecido es necesario para el programa **Antismash**. Pueden revisar la ayuda en linea que se implemento:

```
$ module help programs/antismash-4.1.0

----------- Module Specific Help for 'programs/antismash-4.1.0' ---------------------------

  antiSMASH : Search a genome sequence for secondary metabolite biosynthesis gene clusters.
  Cuidado: se necesita teclear el source despues de cargar el modulo:
  $ module load programs/antismash-4.1.0
  $ source activate antismash
  (antismash) $
   trabajo co antismash...

  Para quitar el ambiente:
  (albacore) $ source deactivate antismash
         $ module unload programs/antismash-4.1.0
         Â¡ Cuidado! antismash asume dos cores por default!
```

El mecanismo parecido deberÃ¡ de ser utilizado para los programas **Bellerophon**, **CompareM FUSTr**, **Qiime2**, y **SuperFocus**

Notes

1. https://github.com/OpenGene/AfterQC

Chapter 3. Databases

3.1. Fasta

Las bases de datos Fasta estan instaladas en **/scratch/DB**. Los archivos aqui presentes son principalmente del sitio ftp de NCBI. Tratamos de realizar las actualisaciones por loe menos 2 veces al semestre.

/ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/ .

- nr.gz | non-redundant protein sequence database with entries from GenPept, Swissprot, PIR, PDF, PDB, and RefSeq
- nt.gz | nucleotide sequence database, with entries from all traditional divisions of GenBank, EMBL, and DDBJ; excluding bulk divisions (gss, sts, pat, est, htg) and wgs entries. Partially non-redundant.
- swissprot.gz | swiss-prot database (last major release)

Se genera tambien una serie de archivos por clados, a partir de nt y nr:

- XX-in-nt.fasta.gz : Secuencias nucleotidos de bacteria, fungi, human, phage o virus desde nt.
- XX-in-nt-50bp.fasta.gz: Misma que la anterior, pero con secuencias >= 50bp
- XX-in-nr.fasta.gz : proteinas de bacteria, fungi, human, phage o virus desde nr.
- XX-in-nr-50bp.fasta.gz : Misma que la anterior, pero con secuencias >= 50aa

3.2. Blast

Formateamos las bases de datos de fasta indicdas en la seccion anterior para Blast. Se encuentran en **/scratch/BlastDB**.

Los modulos del programa blast se encargan de definir de manera automatica la variable **BLASTDB** a este directorio.

Las bases tienen estos nombres , para utilizar com la opcion **-db** de blast:

- nt : nucleotide sequence database.
- nr : non-redundant protein sequence database with entries from GenPept, Swissprot, PIR, PDF, PDB, and RefSeq
- swissprot : swiss-prot database (last major release)
- XX-in-nt-50bp : Secuencias nucleotidos de bacteria, fungi, human o virus , secuencias >= 50bp
- XX-in-nr-50bp : Proteinas de bacteria, fungi, human o virus , secuencias >= 50bp

Por ejemplo, para buscar secuencias de nucleotidos relacionados con bacterias, utilizaria la linea de comando siguiente:

```
$ blastn -db bacteria-in-nt-50bo -query mifastafile ...
```

Chapter 4. Copyrights

4.1. Copyrigths de los programas

Los programas que se presentan en ese documento son libres de derecho de pago.

The software contained in this distribution is released under the academic license agreement which requires to acknowledge the use of the software that results in any published work.