Software list

Below you will find a diverse toolbox comprising numerous tools for the analysis of biological sequences. Not all of the tools will be used in the course, but maybe you are interested in taking a look around to see what is currently around.

Type	Name	Description	GUI	Installation	URL
Read simulation	ART	Illumina, 454 and Solid read simulator	yes	bioconda	Link
Read processing	FastQC	Generates summary statistics and overview information for DNA and RNA seq data	yes	bioconda	Link
	SequelTools	Software package for the quality control, filtering and subread selection of PacBio reads	no	manual	Link
	Trimmomatic	Adapter clipping and quality trimming of short read data	no	bioconda/conda-forge	Link
Transcriptome assembly	Trinity	De novo reconstruction of transcriptomes from RNA-seq data	no	bioconda	Link
Genome assembly	Flye	Fast and accurate de novo assembler for single molecule sequencing reads (suitable for PacBio HiFi reads)	no	bioconda	Link
	CANU	Assembler for High Noise / single molecule sequencing data (Overlap)	no	bioconda	Link
	Velvet	Sequence assembler for short reads (deBruijn graph)	no	bioconda	Link
	SPades	Intended for both standard isolates and single-cell MDA bacteria assemblies (deBruijn Graph)	no	bioconda	Link
	Mira	Whole genome shotgun and EST sequence assembler for Sanger, 454, Solexa (Illumina), IonTorrent data and PacBio (Overlap)	no	bioconda	Link
Database search	NCBI Blast	Heuristic for the rapid identification of significantly similar sequences in a sequence data base using local alignments	no	bioconda	Link
	NCBI Legacy Blast	Heuristic for the rapid identification of significantly similar sequences in a sequence data base using local alignments. These C toolkit binaries are no longer maintained and supported	no	biocore	Link
	Diamond	Accelerated BLAST compatible local sequence aligner	no	bioconda	Link
Gene prediction	maker	A comprehensive pipeline for genome annotation	no	bioconda	Link

Type	Name	Description	GUI	Installation	URL
	funannotate	A pipeline for gene annotation in fungal genomes	no	bioconda	link
	Transdecoder	Identify candidate coding regions within transcript sequences	no	bioconda	Link
Gene set completeness	Busco	Assessing genome assembly and annotation completeness with Benchmarking Universal Single-Copy Orthologs	no	bioconda	Link
	fCAT	Gene set completeness assessment tool using domain-architecture aware targeted ortholog searches	no	pip	link
Repeat annotation	Repeat Masker	Smith-Waterman based identification and optional masking of repeats provided in a repeat database	no	bioconda	Link
Genome visualization	JBrowse Desktop	Desktop version of JBrowse that does not need any web server configuration	yes	manually	Link
	JBrowse	A browser based viewer for genomes and genome-wide annotation	yes	bioconda	Link
	IGV	Integrative Genomics Viewer. Fast, efficient, scalable visualization tool for genomics data and annotations	yes	bioconda	Link
	igvtools	The igvtools utility provides a set of tools for pre-processing data files.	no	bioconda	Link
Structural variant detection	Lumpy	A general probabilistic framework for structural variant discovery	no	bioconda	Link
	Delly	A SV caller leveraging multiple signals	no	bioconda	Link
	Manta	A SV caller leveraging multiple signals	no	bioconda	Link
Structural variation comparison	SURVIOR	A tool kit to compare, merge, and generate stats over SVs vcf files	no	bioconda	Link
SAM/BAM manipulation	Samtools	A tool kit to operate with sam/bam files	no	bioconda	Link
	BCFTOOLS	BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF	no	bioconda	LINK
Read mapping	BWA	A method to align short reads	no	bioconda	Link
	Bowtie2	Ultrafast and memory-efficient tool for aligning sequencing reads to long reference sequences	no	bioconda	Link
SNP caller	xAtlas	A method to detect SNP + indels	no	bioconda	Link

Type	Name	Description	GUI	Installation	URL
RNA-Seq mapping	STAR	A method to align RNA-Seq data	no	bioconda	Link
	Hisat2	A method to align RNA-Seq data	no	bioconda	Link
	Kallisto	A method to align RNA-Seq data	no	bioconda	Link
	TopHat	A spliced read mapper for RNA-Seq	no	bioconda	Link
Assembly evaluation	Quast	Quality Assessment Tool for Genome Assemblies	yes	bioconda	Link
	RnaQuast	Quality evaluation tool for assembled transcripts		bioconda	Link
Taxonomic Assignment	Megan	interactive microbiome analysis tool	no	manually	Link
	Krona Tools	Interaktive viewer for metagenome composition	yes	bioconda	Link
Ortholog search	InParanoid	Pairwise ortholog search tool	no	manually	Link
	OMA	Ortholog matrix project	no	manually	Link
	fDOG	Targeted ortholog search tool	no	pip	Link
Phylogenetic Profiling	PhyloProfile	A browser based tool for visualizing and exploring phylogenetic profiles	yes	Bioconductor	Link
Phylogeny reconstruction	RAxML	Phylogenetics - Randomized Axelerated Maximum Likelihood	no	bioconda	Link
	ProtTest3	ProtTest is a bioinformatic tool for the selection of best-fit models of aminoacid replacement for the data at hand	yes	manually	Link
Tree visualization	FigTree	Graphical viewer of phylogenetic trees	yes	manually	Link
	iTOL	Interactive software for the visualisation and annotation of phylogenetic trees	yes	Web tool	Link
	matt	Interactive tree visualisation, modification and topology testing	yes	pip	Link
Sequence alignment	Muscle	Multiple sequence alignment	no	bioconda	Link

Additional installation information

RepeatMasker

Once you have installed the repeat masker, either directly or via the installation of the maker pipeline, you will need to install the RepBase repeat library. To do so, identify the location of your RepeatMasker installation, download the current release of RepBase repeat library - note, you will have to complete a free registration before you can download the file. Move the archive file name *RepBaseRepeatMaskerEdition-20170127.tar.gz* into the RepeatMasker directory in

```
~/anaconda/envs/compgen/share/RepeatMasker
```

and unpack it by typing

```
tar -xzf RepBaseRepeatMaskerEdition-20170127.tar.gz
```

JBrowse

In case you want to use the desktop version of JBrowse for your work, download the version that matches your operating system from the JBrowse web sites and follow the installation guidelines provided with the software.

Databases

Databases and web tools:

- NCBI - https://www.ncbi.nlm.nih.gov/
- KEGG - http://www.genome.jp/kegg/
- ENSEMBL - http://www.ensembl.org/index.html
- Uniprot - http://www.uniprot.org/
- PFAM - http://pfam.xfam.org/
- InterPro - https://www.ebi.ac.uk/interpro/
- HMMER - http://hmmer.org/

- gNOME - http://ghubs.izn-ffm.intern:5000/g-nom/assemblies/list ❗ This is an ongoing project, and we will be working with an alpha version of this web tool. The URL is reachable only from within our network.

- Back to EcoEvo course
- Back to PBioC course
- Back to DigiKomp course
- Back to MBW course