EDAM ontology and Bio.Tools registry

The ontology and the registry for bioinformatics

Matúš Kalaš

@matuskalas

Computational Biology Unit (CBU), Department of Informatics

University of Bergen, Norway

EDAM

The ontology of bioinformatics data and methods

 @edamontology

 edamontology.org

 /edamontology/edamontology

 http://www.ebi.ac.uk/ols/ontologies/edam

 bioportal.bioontology.org/ontologies/EDAM

- Searching for tools and other resources
  - Tools & data integration
    - Data provenance
      - Text mining

Topic

Phylogenetics

Protein classification

Data

Sequence trace

Position weight matrix

Operation

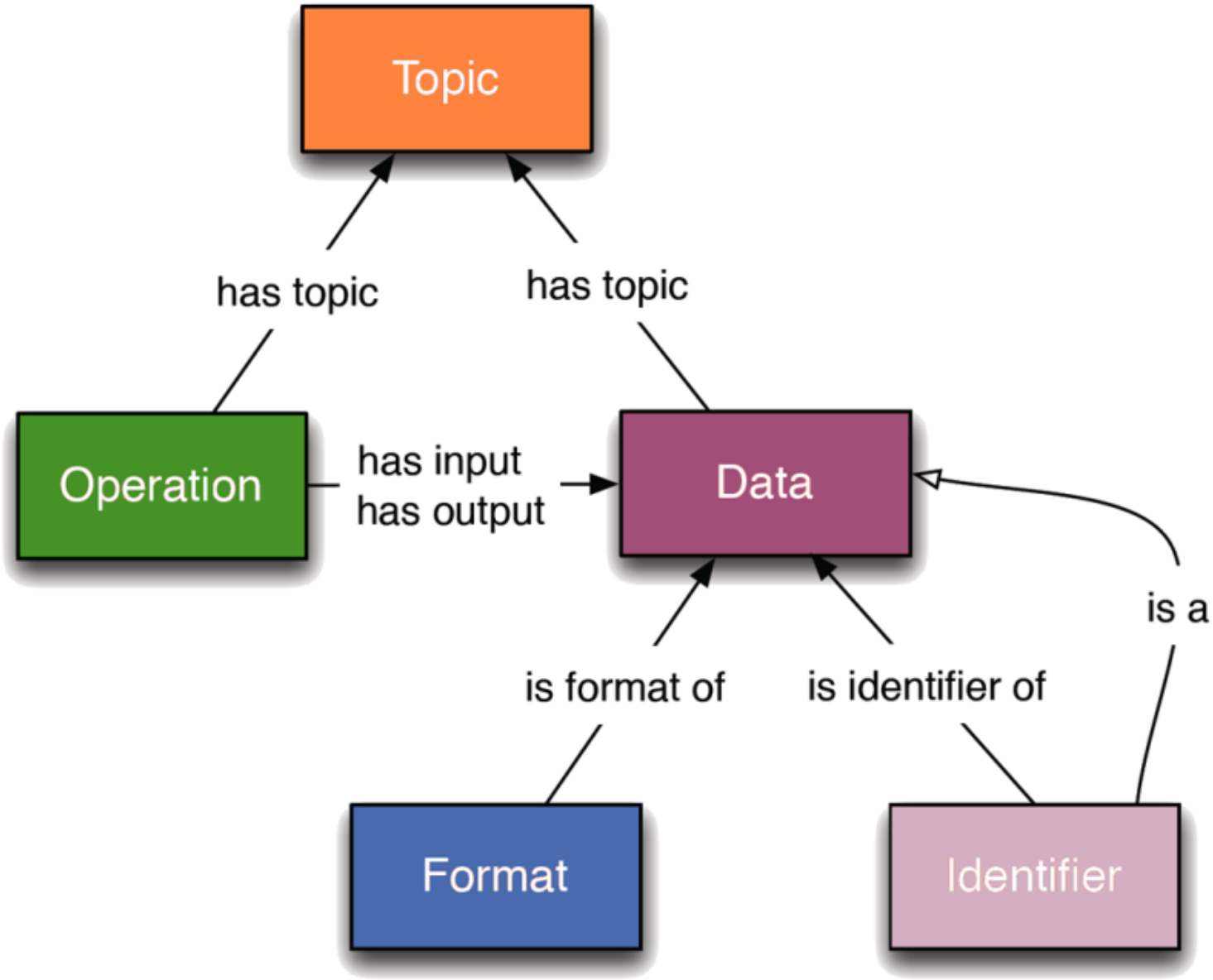
Sequence alignment construction

Molecular dynamics simulation

Format

FASTQ

SBML


Topic
has topic
has topic
Operation
has input
has output
Data
is a
is format of
is identifier of
Format
Identifier

EDAM is used in numerous projects

CCISB Software Catalogue

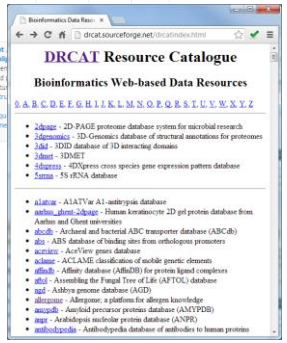
CCP4

instruct Integrating Biology

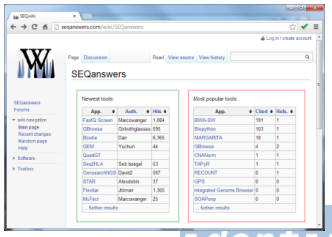
Institut Pasteur

SIB

...



BioXSD



identifiers.org

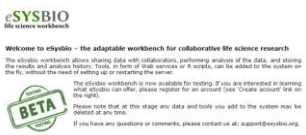
Galaxy

#CommonWL

Bio-jETI.

emboss

The European Molecular Biology Open Software Suite



debian-med

Bio.Tools

The registry of bioinformatics tools and data services

 [@bio_dot_tools](#)

 [https://bio.tools](#)

 [/bio-tools](#)

MBRACE Grid Network of excellence

BioMedBridges

elixir

and other supporting organisations

D38–D47 Nucleic Acids Research, 2016, Vol. 44, Database issue
doi: 10.1093/nar/gkv1116
Published online 3 November 2015

Tools and data services registry: a community effort to document bioinformatics resources

Jon Ison[1,*], Kristoffer Rapacki[1], Hervé Ménager[2], Matúš Kalaš[3], Emil Rydza[1], Piotr Chmura[1], Christian Anthon[4], Niall Beard[5], Karel Berka[6], Dan Bolser[7], Tim Booth[8], Anthony Bretaudeau[9,10], Jan Brezovsky[11], Rita Casadio[12], Gianni Cesareni[13], Frederik Coppens[14,15], Michael Cornell[16], Gianmauro Cuccuru[17], Kristian Davidsen[1], Gianluca Della Vedova[18], Tunca Dogan[19], Olivia Doppelt-Azeroual[2], Laura Emery[7], Elisabeth Gasteiger[20], Thomas Gatter[21], Tatyana Goldberg[22], Marie Grosjean[23], Björn Grüning[24], Manuela Helmer-Citterich[25], Hans Ienasescu[26], Vassilios Ioannidis[20], Martin Closter Jespersen[1], Rafael Jimenez[7], Nick Juty[7], Peter Juvan[27], Maximilian Koch[7], Camille Laibe[7], Jing-Woei Li[28,29], Luana Licata[13], Fabien Mareuil[2], Ivan Mičetić[30], Rune Møllegaard Friborg[31], Sebastien Moretti[20,32], Chris Morris[33], Steffen Möller[34,35], Aleksandra Nenadic[5], Hedi Peterson[36], Giuseppe Profiti[12], Peter Rice[37], Paolo Romano[38], Paola Roncaglia[7], Rabie Saidi[19], Andrea Schafferhans[22], Veit Schwämmle[39], Callum Smith[40], Maria Maddalena Sperotto[1], Heinz Stockinger[20], Radka Svobodová Vařeková[41], Silvio C.E. Tosatto[30], Victor de la Torre[42], Paolo Uva[17], Allegra Via[43], Guy Yachdav[22], Federico Zambelli[44], Gert Vriend[45], Burkhard Rost[22], Helen Parkinson[7], Peter Løngreen[1] and Søren Brunak[1,46]

[1] Center for Biological Sequence Analysis Department of Systems Biology, Technical University of Denmark, Denmark, [2] Centre d'Informatique pour la Biologie, C3BI, Institut Pasteur, France, [3] Computational Biology Unit, Department of Informatics, University of Bergen, Norway, [4] Department of Veterinary Clinical and Animal Sciences, Faculty for Health and Medical Sciences, University of Copenhagen, Denmark, [5] School of Computer Science, University of Manchester, UK, [6] Department of Physical Chemistry, RCPTM, Faculty of Science, Palacky University, Czech Republic, [7] The European Bioinformatics Institute (EMBL-EBI), UK, [8] NEBC Wallingford, Centre for Ecology and Hydrology, UK, [9] INRA, UMR Institut de Génétique, Environnement et Protection des Plantes (IGEPP), BioInformatics Platform for Agroecosystems Arthropods (BIPAA), France, [10] INRIA, IRISA, GenOuest Core Facility, France, [11] Loschmidt Laboratories, Department of Experimental Biology and Research Centre for Toxic Compounds in the Environment RECETOX, Masaryk University, Czech Republic, [12] Bologna Biocomputing Group, University of Bologna, Italy, [13] Dept. of Biology, University of Rome Tor Vergata, Italy, [14] Department of Plant Systems Biology, VIB, Belgium, [15] Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium, [16] Faculty of Life Sciences, University [illegible] of, UK, [illegible], [17] [illegible], [18] [illegible], [19] [illegible], [20] [illegible], Swiss Institute of Bioinformatics, Switzerland, [21] Faculty of Technology and Center for Biotechnology, University Bielefeld, Germany, [22] Department of Informatics, Bioinformatics-I12, TUM, Germany, [23] Institut Français de Bioinformatique (French Institute of Bioinformatics), CNRS, UMS3601, France, [24] Albert-Ludwigs-Universität Freiburg, Fahnenbergplatz, 79085 Freiburg, [25] Centre for Molecular Bioinformatics, Dept. of Biology, University of Rome Tor Vergata, Italy, [26] Bioinformatics Centre, Department of Biology, University of Copenhagen, Denmark, [27] Centre for Functional Genomics and Biochips, Faculty of Medicine, University of Ljubljana, Slovenia, [28] Faculty of Medicine, The Chinese University of Hong Kong, China, [29] Hong Kong Bioinformatics Centre, School of Life Sciences, The Chinese [illegible] Department of Biomedical Sciences, University of Padua, Italy, [31] Bioinformatics

The use cases of the Bio.Tools registry are …

find

understand

compare

select

attribute

…

… bioinformatics tools and their usage

Bio.Tools should provide answers to various questions

I have this task in mind, what tools are there for it?

I need a tool that reads or writes data in a specific format?

What tools are available in a general area, *e.g.* proteomics?

What are the interfaces to a given tool?

What are the implementations of a particular algorithm?

What are the parameters of this tool and how do I use it?

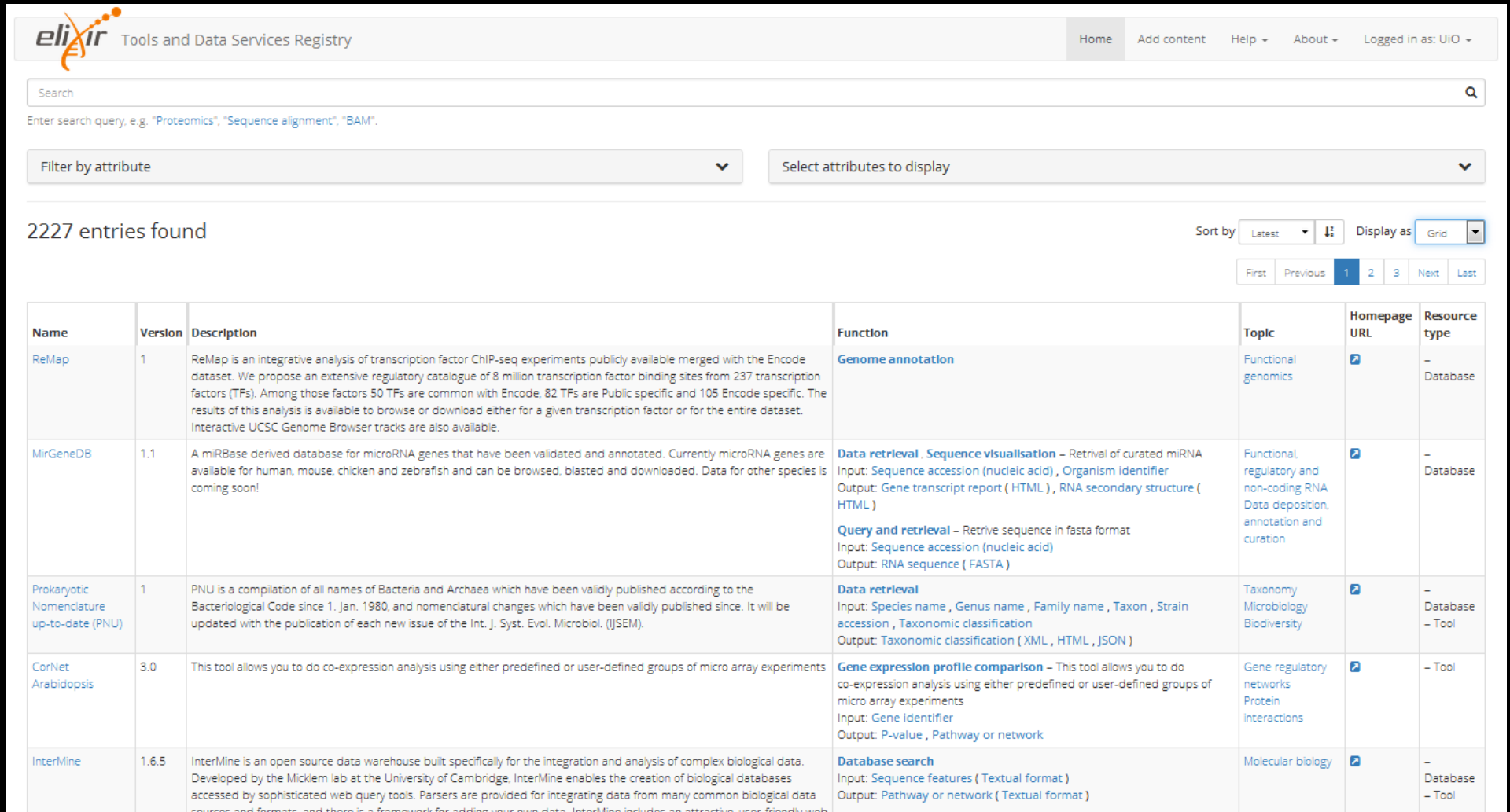
What is the last stable version of a tool?

What are the outputs of a particular research institute or infrastructure?

What resources arose from a particular grant?

and many others …

https://bio.tools



elixir Tools and Data Services Registry

Home | Add content | Help | About | Logged in as: UiO

Search

Enter search query, e.g. "Proteomics", "Sequence alignment", "BAM".

Filter by attribute

Select attributes to display

2227 entries found

Sort by Latest | Display as Grid

First | Previous | 1 | 2 | 3 | Next | Last

Name	Version	Description	Function	Topic	Homepage URL	Resource type
ReMap	1	ReMap is an integrative analysis of transcription factor ChIP-seq experiments publicly available merged with the Encode dataset. We propose an extensive regulatory catalogue of 8 million transcription factor binding sites from 237 transcription factors (TFs). Among those factors 50 TFs are common with Encode, 82 TFs are Public specific and 105 Encode specific. The results of this analysis is available to browse or download either for a given transcription factor or for the entire dataset. Interactive UCSC Genome Browser tracks are also available.	**Genome annotation**	Functional genomics		– Database
MirGeneDB	1.1	A miRBase derived database for microRNA genes that have been validated and annotated. Currently microRNA genes are available for human, mouse, chicken and zebrafish and can be browsed, blasted and downloaded. Data for other species is coming soon!	**Data retrieval** , **Sequence visualisation** – Retrival of curated miRNA Input: Sequence accession (nucleic acid) , Organism identifier Output: Gene transcript report (HTML) , RNA secondary structure (HTML) **Query and retrieval** – Retrive sequence in fasta format Input: Sequence accession (nucleic acid) Output: RNA sequence (FASTA)	Functional, regulatory and non-coding RNA Data deposition, annotation and curation		– Database
Prokaryotic Nomenclature up-to-date (PNU)	1	PNU is a compilation of all names of Bacteria and Archaea which have been validly published according to the Bacteriological Code since 1. Jan. 1980, and nomenclatural changes which have been validly published since. It will be updated with the publication of each new issue of the Int. J. Syst. Evol. Microbiol. (IJSEM).	**Data retrieval** Input: Species name , Genus name , Family name , Taxon , Strain accession , Taxonomic classification Output: Taxonomic classification (XML , HTML , JSON)	Taxonomy Microbiology Biodiversity		– Database – Tool
CorNet Arabidopsis	3.0	This tool allows you to do co-expression analysis using either predefined or user-defined groups of micro array experiments	**Gene expression profile comparison** – This tool allows you to do co-expression analysis using either predefined or user-defined groups of micro array experiments Input: Gene identifier Output: P-value , Pathway or network	Gene regulatory networks Protein interactions		– Tool
InterMine	1.6.5	InterMine is an open source data warehouse built specifically for the integration and analysis of complex biological data. Developed by the Micklem lab at the University of Cambridge, InterMine enables the creation of biological databases accessed by sophisticated web query tools. Parsers are provided for integrating data from many common biological data	**Database search** Input: Sequence features (Textual format) Output: Pathway or network (Textual format)	Molecular biology		– Database – Tool

https://bio.tools

2227 entries found

Sort by Latest | Display as Cards

First Previous 1 2 3 Next Last

ReMap

http://tagc.univ-mrs.fr/remap/index.php

ReMap is an integrative analysis of transcription factor ChIP-seq experiments publicly available merged with the Encode dataset. We propose an extensive regulatory catalogue of 8 million transcription factor binding sites from 237 transcription factors (TFs). Among those factors 50 TFs are common with Encode, 82 TFs are Public specific and 105 Encode specific. The results of this analysis is available to browse or download either for a given transcription factor or for the entire dataset. Interactive UCSC Genome Browser tracks are also available.

Genome annotation

Addition date about 21 hours ago
Affiliation inserm.fr
Topic Functional genomics
Resource Type Database
Interface Web UI

Documentation

Read the docs

Download

Citation instructions

MirGeneDB v.1.1 Edit

http://mirgenedb.org/

A miRBase derived database for microRNA genes that have been validated and annotated. Currently microRNA genes are available for human, mouse, chicken and zebrafish and can be browsed, blasted and downloaded. Data for other species is coming soon!

Data retrieval, Sequence visualisation

Retrival of curated miRNA

Inputs: Sequence accession (nucleic acid) , Organism identifier

Outputs: Gene transcript report (HTML) , RNA secondary structure (HTML)

Query and retrieval

Retrive sequence in fasta format

Inputs: Sequence accession (nucleic acid)

Outputs: RNA sequence (FASTA)

Addition date a day ago
Affiliation UiO
Topic Functional, regulatory and non-coding RNA, Data deposition, annotation and curation
Resource Type Database
Interface Web UI
Collection UiO tools

Documentation

Read the docs

Prokaryotic Nomenclature up-to-date (PNU)

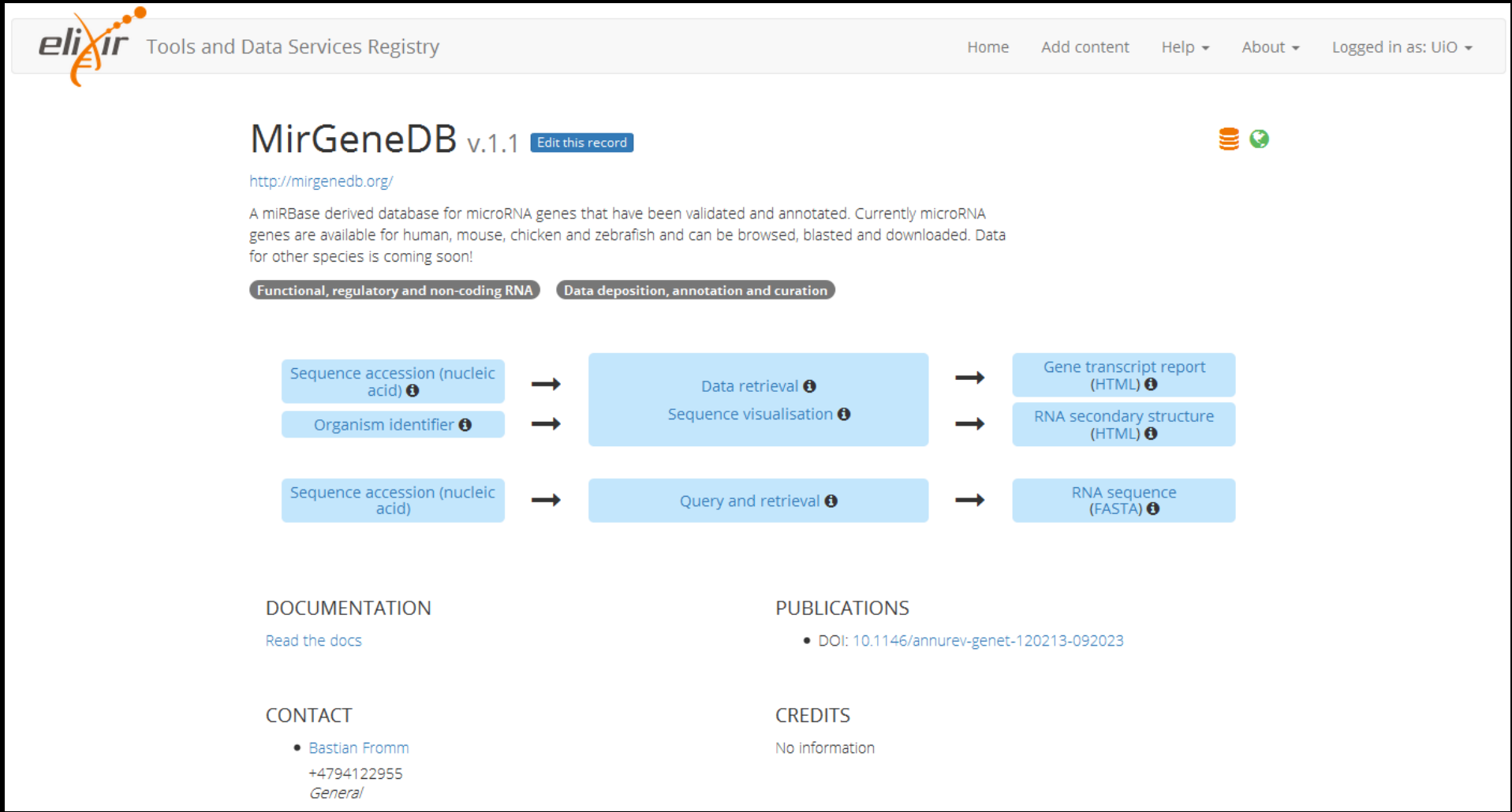
http://bacdive.dsmz.de/api/pnu/

PNU is a compilation of all names of Bacteria and Archaea which have been validly published according to the Bacteriological Code since 1. Jan. 1980, and nomenclatural changes which have been validly published since. It will be updated with the publication of each new issue of the Int. J. Syst. Evol. Microbiol. (IJSEM).

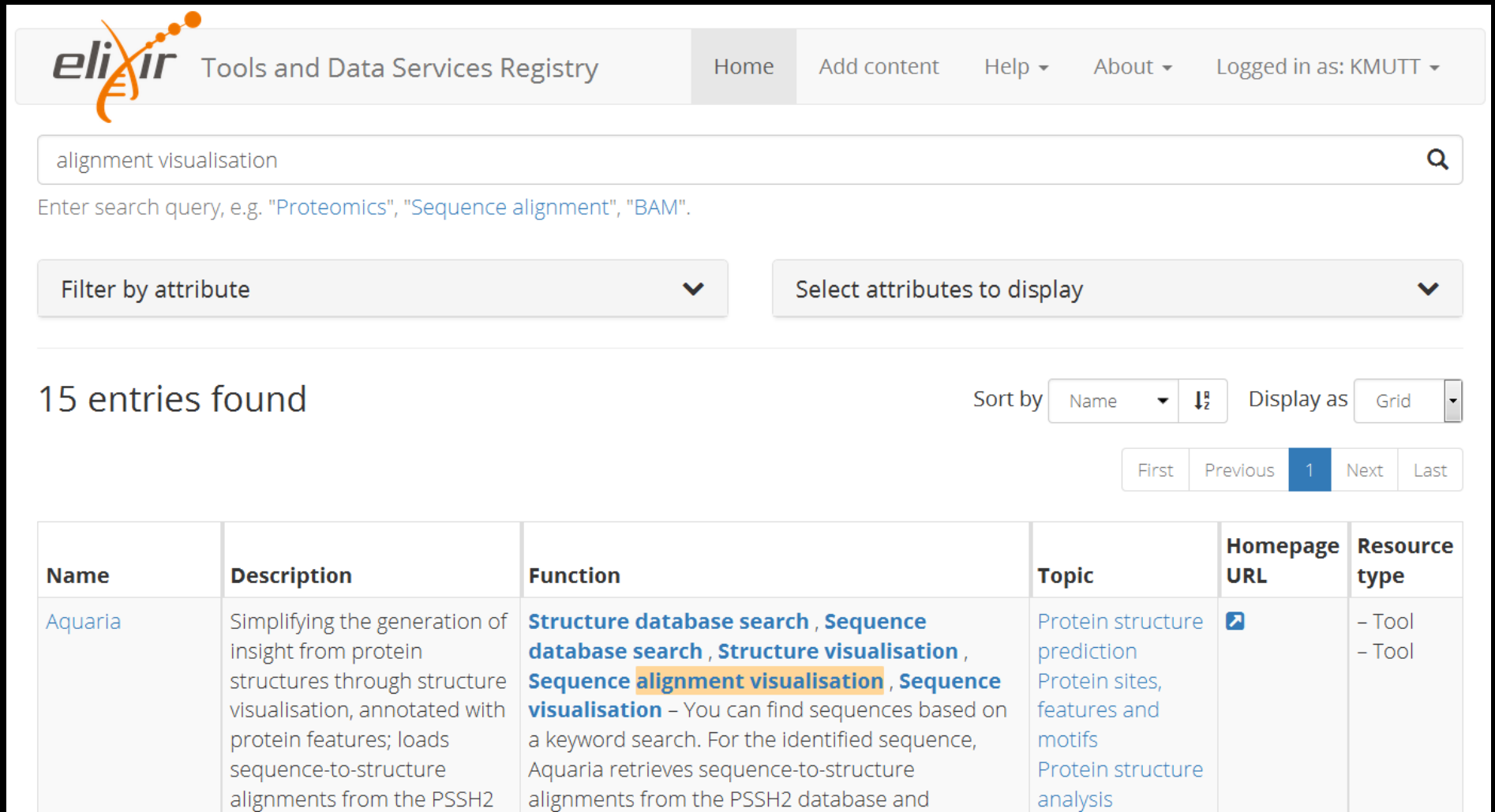
Addition date 6 days ago
Affiliation bacdive
Topic Taxonomy, Microbiology, Biodiversity
Resource Type Database, Tool
Interface API
Maturity Stable
Cost Free

Documentation

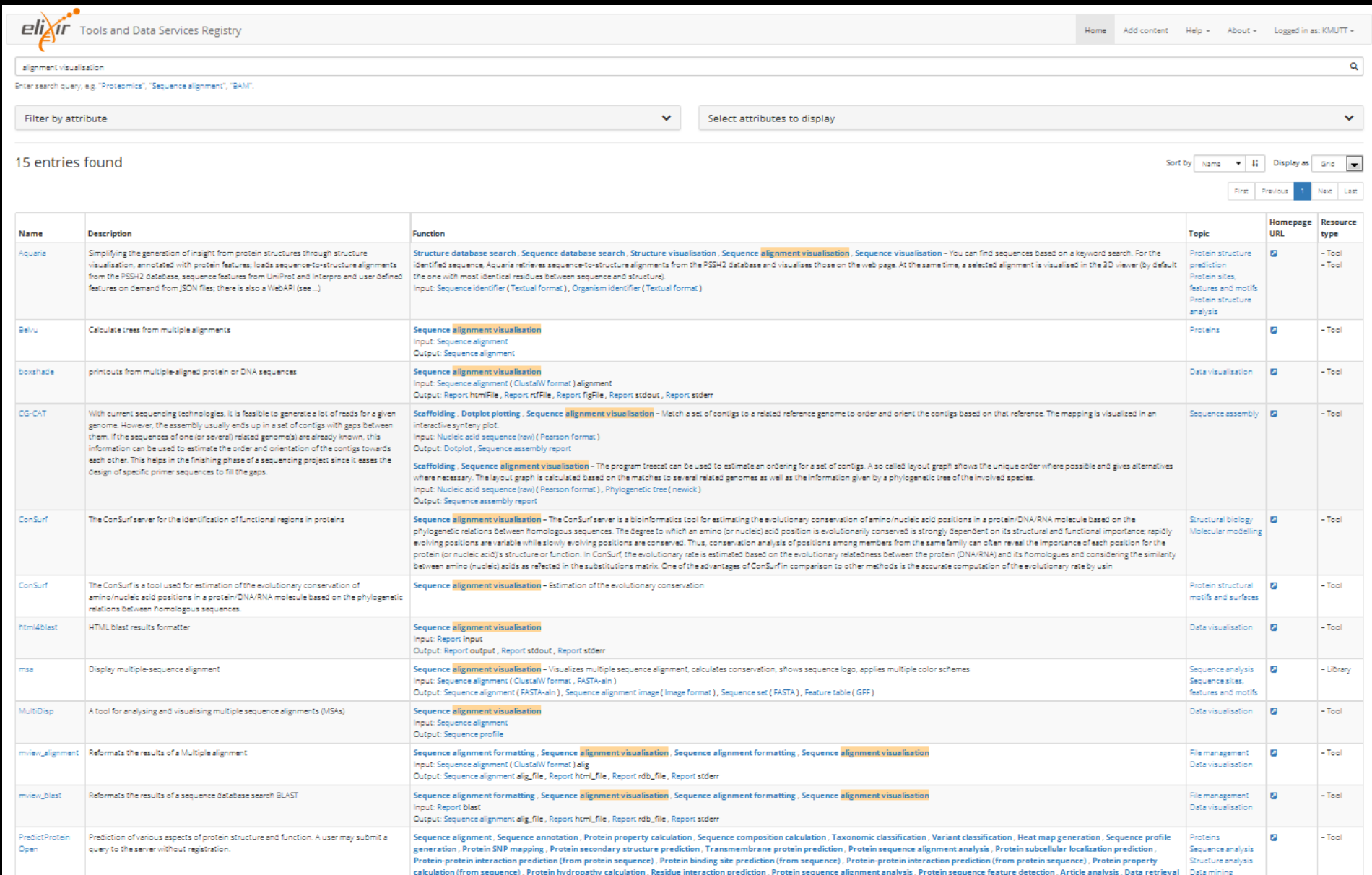
Example record: MirGeneDB



https://bio.tools/?q=alignment%20visualisation



https://bio.tools/?q=alignment%20visualisation



Tools and Data Services Registry

Home | Add content | Help | About | Logged in as: KMUTT

alignment visualisation

Enter search query, e.g. "Proteomics", "Sequence alignment", "BAM".

Filter by attribute

Select attributes to display

15 entries found

Sort by: Name | Display as: Grid

First | Previous | 1 | Next | Last

Name	Description	Function	Topic	Homepage URL	Resource type
Aquaria	Simplifying the generation of insight from protein structures through structure visualisation, annotated with protein features; loads sequence-to-structure alignments from the PSSH2 database, sequence features from UniProt and Interpro and user defined features on demand from JSON files; there is also a WebAPI (see ...)	Structure database search, Sequence database search, Structure visualisation, Sequence alignment visualisation, Sequence visualisation – You can find sequences based on a keyword search. For the identified sequence, Aquaria retrieves sequence-to-structure alignments from the PSSH2 database and visualises those on the web page. At the same time, a selected alignment is visualised in the 3D viewer (by default the one with most identical residues between sequence and structure). Input: Sequence identifier (Textual format), Organism identifier (Textual format)	Protein structure prediction Protein sites, features and motifs Protein structure analysis		– Tool – Tool
Belvu	Calculate trees from multiple alignments	Sequence alignment visualisation Input: Sequence alignment Output: Sequence alignment	Proteins		– Tool
boxshade	printouts from multiple-aligned protein or DNA sequences	Sequence alignment visualisation Input: Sequence alignment (ClustalW format) alignment Output: Report htmlFile, Report rtfFile, Report figFile, Report stdout, Report stderr	Data visualisation		– Tool
CG-CAT	With current sequencing technologies, it is feasible to generate a lot of reads for a given genome. However, the assembly usually ends up in a set of contigs with gaps between them. If the sequences of one (or several) related genome(s) are already known, this information can be used to estimate the order and orientation of the contigs towards each other. This helps in the finishing phase of a sequencing project since it eases the design of specific primer sequences to fill the gaps.	Scaffolding, Dotplot plotting, Sequence alignment visualisation – Match a set of contigs to a related reference genome to order and orient the contigs based on that reference. The mapping is visualized in an interactive synteny plot. Input: Nucleic acid sequence (raw) (Pearson format) Output: Dotplot, Sequence assembly report Scaffolding, Sequence alignment visualisation – The program treecat can be used to estimate an ordering for a set of contigs. A so called layout graph shows the unique order where possible and gives alternatives where necessary. The layout graph is calculated based on the matches to several related genomes as well as the information given by a phylogenetic tree of the involved species. Input: Nucleic acid sequence (raw) (Pearson format), Phylogenetic tree (newick) Output: Sequence assembly report	Sequence assembly		– Tool
ConSurf	The ConSurf server for the identification of functional regions in proteins	Sequence alignment visualisation – The ConSurf server is a bioinformatics tool for estimating the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on the phylogenetic relations between homologous sequences. The degree to which an amino (or nucleic) acid position is evolutionarily conserved is strongly dependent on its structural and functional importance: rapidly evolving positions are variable while slowly evolving positions are conserved. Thus, conservation analysis of positions among members from the same family can often reveal the importance of each position for the protein (or nucleic acid)'s structure or function. In ConSurf, the evolutionary rate is estimated based on the evolutionary relatedness between the protein (DNA/RNA) and its homologues and considering the similarity between amino (nucleic) acids as reflected in the substitutions matrix. One of the advantages of ConSurf in comparison to other methods is the accurate computation of the evolutionary rate by usin	Structural biology Molecular modelling		– Tool
ConSurf	The ConSurf is a tool used for estimation of the evolutionary conservation of amino/nucleic acid positions in a protein/DNA/RNA molecule based on the phylogenetic relations between homologous sequences.	Sequence alignment visualisation – Estimation of the evolutionary conservation	Protein structural motifs and surfaces		– Tool
html4blast	HTML blast results formatter	Sequence alignment visualisation Input: Report input Output: Report output, Report stdout, Report stderr	Data visualisation		– Tool
msa	Display multiple-sequence alignment	Sequence alignment visualisation – Visualizes multiple sequence alignment, calculates conservation, shows sequence logo, applies multiple color schemes Input: Sequence alignment (ClustalW format, FASTA-aln) Output: Sequence alignment (FASTA-aln), Sequence alignment image (Image format), Sequence set (FASTA), Feature table (GFF)	Sequence analysis Sequence sites, features and motifs		– Library
MultiDisp	A tool for analysing and visualising multiple sequence alignments (MSAs)	Sequence alignment visualisation Input: Sequence alignment Output: Sequence profile	Data visualisation		– Tool
mview_alignment	Reformats the results of a Multiple alignment	Sequence alignment formatting, Sequence alignment visualisation, Sequence alignment formatting, Sequence alignment visualisation Input: Sequence alignment (ClustalW format) alig Output: Sequence alignment alig_file, Report html_file, Report rdb_file, Report stderr	File management Data visualisation		– Tool
mview_blast	Reformats the results of a sequence database search BLAST	Sequence alignment formatting, Sequence alignment visualisation, Sequence alignment formatting, Sequence alignment visualisation Input: Report blast Output: Sequence alignment alig_file, Report html_file, Report rdb_file, Report stderr	File management Data visualisation		– Tool
PredictProtein Open	Prediction of various aspects of protein structure and function. A user may submit a query to the server without registration.	Sequence alignment, Sequence annotation, Protein property calculation, Sequence composition calculation, Taxonomic classification, Variant classification, Heat map generation, Sequence profile generation, Protein SNP mapping, Protein secondary structure prediction, Transmembrane protein prediction, Protein sequence alignment analysis, Protein subcellular localization prediction, Protein-protein interaction prediction (from protein sequence), Protein binding site prediction (from sequence), Protein-protein interaction prediction (from protein sequence), Protein property calculation (from sequence), Protein hydropathy calculation, Residue interaction prediction, Protein sequence alignment analysis, Protein sequence feature detection, Article analysis, Data retrieval	Proteins Sequence analysis Structure analysis Data mining		– Tool

Thank you for your attention!

Bio.Tools


@bio_dot_tools


https://bio.tools


/bio-tools

EDAM


@edamontology


edamontology.org


/edamontology/edamontology


www.ebi.ac.uk/ols/ontologies/edam


bioportal.bioontology.org/ontologies/EDAM