Resource Summary Report

Generated by NIF on Sep 11, 2026

MedBlast

RRID:SCR_008202
Type: Tool

Proper Citation

MedBlast (RRID:SCR_008202)

Resource Information

URL: http://medblast.sibsnet.org/

Proper Citation: MedBlast (RRID:SCR_008202)

Description: THIS RESOURCE IS NO LONGER IN SERVICE, documented August 29, 2016. An algorithm that finds articles most relevant to a genetic sequence. In the genomic era, researchers often want to know more information about a biological sequence by retrieving its related articles. However, there is no available tool yet to achieve conveniently this goal. Here, a new literature-mining tool MedBlast is developed, which uses natural language processing techniques, to retrieve the related articles of a given sequence. An online server of this program is also provided. The genome sequencing projects generate such a large amount of data every day that many molecular biologists often encounter some sequences that they know nothing about. Literature is usually the principal resource of such information. It is relatively easy to mine the articles cited by the sequence annotation; however, it is a difficult task to retrieve those relevant articles without direct citation relationship. The related articles are those described in the given sequence (gene/protein), or its redundant sequences, or the close homologs in various species. They can be divided into two classes: direct references, which include those either cited by the sequence annotation or citing the sequence in its text; indirect references, those which contain gene symbols of the given sequence. A few additional issues make the task even more complicated: (1) symbols may have aliases; and (2) one sequence may have a couple of relatives that we want to take into account too, which include redundant (e.g. protein and gene sequences) and close homologs. Here the issues are addressed by the development of the software MedBlast, which can retrieve the related articles of the given sequence automatically. MedBlast uses BLAST to extend homology relationships, precompiled species-specific thesauruses, a useful semantics technique in natural language processing (NLP), to extend alias relationship, and EUtilities toolset to search and retrieve corresponding articles of each sequence from PubMed. MedBlast take a sequence in FASTA format as input. The program first uses BLAST to search the GenBank nucleic acid and protein non-redundant (nr) databases, to extend to those homologous and corresponding nucleic acid and protein sequences. Users can input the BLAST results directly, but it is recommended to input the result of both protein

and nucleic acid nr databases. The hits with low e-values are chosen as the relatives because the low similarity hits often do not contain specific information. Very long sequences, e.g. 100k, which are usually genomic sequences, are discarded too, for they do not contain specific direct references. User can adjust these parameters to meet their own needs.

Synonyms: MedBlast

Resource Type: software resource

Keywords: gene, article, biological, data, genome, genomic, homolog, literature, medline interfaces, mining, molecular, protein, sequence, specie

Funding: National Natural Science Foundation of China 39990600-03; Knowledge Innovation Program of the Chinese Academy of Sciences KSCX2-2-07; Knowledge Innovation Program of the Chinese Academy of Sciences KJCX1-08

Availability: THIS RESOURCE IS NO LONGER IN SERVICE

Resource Name: MedBlast

Resource ID: SCR_008202

Alternate IDs: nif-0000-21253

Record Creation Time: 20220129T080246+0000

Record Last Update: 20260905T132619+0000

Ratings and Alerts

No rating or validation information has been found for MedBlast.

No alerts have been found for MedBlast.

Data and Source Information

Source: SciCrunch Registry

Usage and Citation Metrics

We found 1 mentions in open access literature.

Listed below are recent publications. The full list is available at NIF .

Krallinger M, et al. (2005) Text-mining and information-retrieval services for molecular biology. Genome biology, 6(7), 224.