BLASTphp: a PHP wrapper for NCBI BLAST API

Ashok Kumar T1, Rajagopal B2
1 Department of Bioinformatics, Noorul Islam College of Arts and Science, Thuckalay - 629180, Tamil Nadu, India
2 Department of Zoology, Government Arts College, Dharmapuri - 636705, Tamil Nadu, India

ABSTRACT

NCBI BLAST is a most popular bioinformatics framework for finding local similarity between two or more biological sequences. It provides integrative access to various biological databases through web user interface, command-line user interface, and application programming interface. BLASTphp is a lightweight PHP library that wraps NCBI BLAST’s RESTful API into a custom graphical user interface BLAST. The BLASTphp program remotely executes the NCBI BLAST through the HTTP/HTTPS interface and elicits the response in HTML, XML, XML2, JSON2, Text, or Tabular (plain) format. However, it can be easily connected to a webserver and integrated with any SQL database which contain the biological sequences. BLASTphp is freely available under GNU General Public License version 3 (GPLv3), at https://github.com/AshokHub/BLASTphp.

Copyright © 2017 International Journal for Computational Biology, http://www.ijcb.in. All rights reserved.

1. INTRODUCTION

BLAST (Basic Local Alignment Search Tool) is a sequence analysis tool widely used for searching homologous sequences in databases by measuring the maximal segment pair (MSP) for two sequences. The BLAST algorithm and its supporting programs were originally developed at the National Institutes of Health (NIH) by Steve Altschul and his team members in 1990, and it was hosted by the National Center for Biotechnology Information (NCBI) [1-3]. It uses a rapid heuristic approach to reduce time-consuming for sequence alignment, but less accurate than the dynamic programming method such as Smith-Waterman algorithm [4]. Modern BLAST consists of many different types of programs for sequence search, namely BLASTN (nucleotide to nucleotide search), TBLASTN (protein to translated nucleotide search), BLASTP (protein to protein search), BLASTX (translated nucleotide to protein search), MegaBLAST (nucleotide to nucleotide search in closely related species), PSI-BLAST (protein to protein search from PSSM), PHI-BLAST (protein with pattern to protein search), RPSBLAST (Reverse-Position-Specific BLAST: protein to PSSM search), and DELTA-BLAST (PSSM from protein to PSSM search using fast RPSBLAST) [5]. Due to the popularity of NCBI BLAST, huge flavours of BLAST programs namely BLASTZ, CDART, GEOBLAST, IgBLAST, BLAT, SNP BLAST, WU BLAST, AB BLAST, DIAMOND, CaBLAST, Paracel BLAST, MIPBlas, etc. were released with different features [6].

NCBI BLAST offers a variety of methods to perform a sequence search, namely Online BLAST, BLAST URL API, Stand-alone BLAST (BLAST+), Remote BLAST+, Cloud BLAST, and C++ BLAST API [5]. The complete source code and detailed documentation of NCBI C++ ToolKit is freely available at FTP site of NCBI and GitHub repository [5]. Because of the performance of NCBI BLAST, some of the open source programming modules/packages such as BioPerl (Bio::Tools::Run::RemoteBlast – Perl module) [7], BioPython
(Bio.Blast.NCBIWWW – Python module) [8], BioJava (org.biojavax.bio.alignment.blast.RemoteQBLastService – Java package) [9], and BioRuby (Bio::Blast::Remote – Ruby module) [10] have been used to wrap BLAST URL API for remote execution of the NCBI BLAST through HTTP interface.

The NCBI BLAST RESTful service is also available in different programming languages from other biological databases such as RCSB PDB, DDBJ, UniProtKB/SwissProt, EMBL EBI, etc. Some sequence analysis and molecular modeling tools such as SWISS MODEL, Swiss-PdbViewer, BIOVIA Discovery Studio, etc. either use the NCBI BLAST RESTful service to retrieve sequences from the database, or standalone command-line BLAST+ to compare sequence with the large set of sequences (local database). BLASTphp is wrapper of NCBI BLAST URL API written in the PHP programming language, and has been distributed on the GitHub open source repository.

2. METHODS

The BLASTphp program is implemented in PHP language and provides a wrapper to the NCBI BLAST URL API. The BLASTphp is a small program and does not require any additional configurations or modules to be included. A simple PHP supporting user interactive web form or a command-line interface is sufficient for remote execution of NCBI BLAST server through BLASTphp program (Figure 1).

The BLASTphp program gets the URL encoded FASTA sequence or NCBI database accession number/gi, and BLAST program parameters such as CMD, DATABASE, PROGRAM, QUERY, etc. from the client interface and passes to the BLAST RESTful API through HTTP/HTTPS interface. The BLAST RESTful API acts as a gateway between BLASTphp and the BLAST server during the execution. Before server execution, it returns the BLAST search request identifier (RID) and request time of execution (RTOE) for the submitted sequence. BLASTphp waits for BLAST server execution till RTOE (in seconds) and elicits the status and/or result in HTML, Text, XML, XML2, JSON2, or Tabular (plain) format to the user client interface. By default, the BLAST output will be in HTML format. The response may also consist of Informational or Status (WAITING|FAILED|UNKNOWN|READY) commands, which are used to track status of the BLAST execution.

2.1. Error handling

If a problem occurs during remote BLAST server execution, NCBI BLAST RESTful service will return a short error message describing the reason for failure of execution. The common reason for the failures are: invalid characters/empty lines in the FASTA sequence/illegal format given in the input, no homologous sequence match found, the BLAST program parameters are invalid, the requested query sequence was too large and requires more time to complete BLAST execution (the standard NCBI BLAST execution time limit is 30 seconds), etc. In technical side, the following HTTP status codes with a message (400 - syntax error in the URL, POST body, etc., 404 - invalid RID; 405 - invalid MIME type in the HTTP Accept header, 504 - request timed out due to server overload, 501 - invalid BLAST program parameters, 500 - server side problems such as a database server down, etc.) will be returned.
3. RESULTS

BLASTphp is a lightweight PHP program which consumes less bandwidth and resource on the web server. Currently, BLASTphp is ready for development of PHP based applications. BLASTphp provides similar functionalities of the native NCBI BLAST server. The graphical web interactive BLAST result of the sequence alignment can be obtained just by including the NCBI server path of the images and javascripts on top of the PHP code. The BLASTphp program can be also be used to execute cloud BLAST servers hosted on Amazon Web Services (AWS), Google Compute Engine (GCE), and Microsoft Azure.

Conflict of interest: None declared.

REFERENCES


