

AIM-BLAST-AJAX Interfaced Multisequence Blast

G. Aravindhan¹, R. Sathish Kumar², K. Subha¹, T.K.Subazini¹, Alpana Dey³, Krishna Kant³ and G. Ramesh Kumar¹

¹Bioinformatics Division, AU-KBC Research Centre, MIT Campus, Anna University, Chennai—600 044, India. ²NRCFOSS, AU-KBC Research Centre, MIT Campus, Anna University, Chennai—600 044, India. ³Ministry of Communications and Information Technology, DT, New Delhi-110 003, India.

Abstract: AIM-BLAST, AJAX Interfaced Multisequence Blast, is a simplified tool developed to facilitate the multiple sequences blast using AJAX as an interface. This tool has been integrated with the SOAP services of EBI NCBI Blast and the functionality of AJAX (Asynchronous Javascript and XML), so as to minimize the enormous bandwidth consumption while carrying out blast analysis for many sequences at an instance. Although a few tools for multiple sequences blast are already available online, they are restricted only to a limited number of genomes and consume several bytes of data transfer for receiving the results. Further, AIM-BLAST also has enhanced features for automated parsing of the Blast results of individual sequence and presenting them as “one sequence-one function” manner. This will save the users time and effort in interpreting the bulky blast results to identify one suitable hit. The results of the blast search in this tool are displayed in an easily interpretable table format that makes the tool user-friendly too. Hence this tool, with a laconic framework, will remain a well structured, flexible and a highly controlled Blast Program for investigating numerous sequences at a stretch with the consumption of reduced level of data transfer.

Availability: AIM-BLAST is freely available at <http://biotool.nrcfosshelpline.in/aimblast/>

Keywords: AIM-BLAST, AJAX, Blast, JavaScript

Introduction

Past decade have created enormous volumes of biological data that are deposited in the online repositories.¹ With such a largely mounted data, it has become the most vital challenge for the scientific community to investigate these raw sequences and reproduce their functions effectively. Delineating such meaningful information will facilitate a better insight into the complex biological systems. In spite of various advanced strategies for identifying the protein functions were carried out earlier, only 50%–60% of genes have been identified with known functions in most completely sequenced genomes.² Therefore, determining the role of proteins become the most focused research areas of post-genome era. Although the traditional biochemical/molecular approaches can produce accurate information, they consume a lot of materials, manpower and man-hour making the process cost ineffective.³

This demands the assistance of Bioinformatics systems to carryout sequence analysis. Of most Bioinformatics approaches, the discovery of sequence homology to a know protein or family of proteins often provides the first clues about the function of a newly sequenced gene. This makes the analyses of the biological sequences using sequence similarity search tools like BLAST,⁴ preliminary but essential step in the Bioinformatics research. BLAST, Basic Local Alignment Search Tool, is one of the most favorite and widely used Bioinformatics program for identifying the similarity between the biological sequences based on several parameters. Blast program is available from different sources including the BLAST utility maintained by EBI, European Bioinformatics Institute, [<http://www.ebi.ac.uk/Tools/blastall>] and the BLAST services offered by the NCBI, National Centre for Biotechnology Information, [<http://blast.ncbi.nlm.nih.gov/Blast.cgi>].

But, these tools are computationally intensive and time consuming as they employ a voluminous amount of data transfer for every analysis. Analyzing a single sequence against a regular Blast program [<http://www.ebi.ac.uk/Tools/blast/>], will itself generate large amount of results in terms of hits accompanied with varied parameters such as E-value, Percentage of Identity, Percentage of Similarity, Blast

Correspondence: G. Ramesh Kumar, Bioinformatics Division, AU-KBC Research Centre, MIT Campus, Anna University, Chennai-600 044, India. Email: gramesh@au-kbc.org G. Aravindhan, Centre for Molecular Simulations, Faculty of ICT, Swinburne University of Technology, Victoria-3122, Australia. Email: aganesan@groupwise.swin.edu.au



Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: <http://creativecommons.org/licenses/by/3.0/>.

score and sequence length. Thus, a lot of human interventions are required in interpreting such huge results and choose one best hit.

Further, these tools utilize the “client pull” also called “meta refresh” approach, in which a query is submitted, the program will forward the sequence to the corresponding server for the analysis to take place and divert the user’s browser to a temporary page with a job ID assigned for the submitted sequence. This temporary browser (Fig. 1) keeps on refreshing until the result is ready in the server and thus, consumes a lot of bandwidth for every single refresh. Also, during the page refresh, the users are forced to sit idle and stare at the refreshing window that creates an unpleasant user experience.

Moreover, these popular Blast tools do not offer services to carryout the multiple sequences blast at an instance. But still, there are a very few tools that are available online for carrying out the analysis for the multiple sequences such as Bblast, [gopher://megasun.bch.umontreal.ca:70/11/CMB/Databases/Blast/bblast] and Blast services offered by National Microbial Pathogen Data Resource⁵ [http://beta.nmpdr.org//cur/FIG/SearchSkeleton.cgi?Class=BlastSearch]. However, these programs limit the blast search against only a restricted number of genome databases but not

against all the genome databases. Hence, there is a pressing need to develop an advanced computational program that will balance all these limitations and handle the sequence annotation better.

Here, we present AIM-BLAST, AJAX Interfaced Multisequence Blast, as an enhanced Blast tool that can potentially handle multiple protein sequences at a stretch, besides consuming a very limited bandwidth. The automated parsing of the results in this tool will help to recognize the significant hit for every sequence and thus, making the tool quicker and user loving than any other heuristic algorithms.

Materials and Methods

AIM-BLAST is a system that has been developed to fulfill the demerits that exist in other blast programs. This program is exclusively made to facilitate multiple sequences blast at an instance against all the genome databases. This program is designed with an efficient process scheme (Fig. 2) using the services offered by the EBI. The application design is written using HTML/Javascripts whereas, the server end of the tool is written using the Perl scripts. Moreover, AJAX⁶ is deployed in AIM-BLAST that serves as an interface between the users and the application.

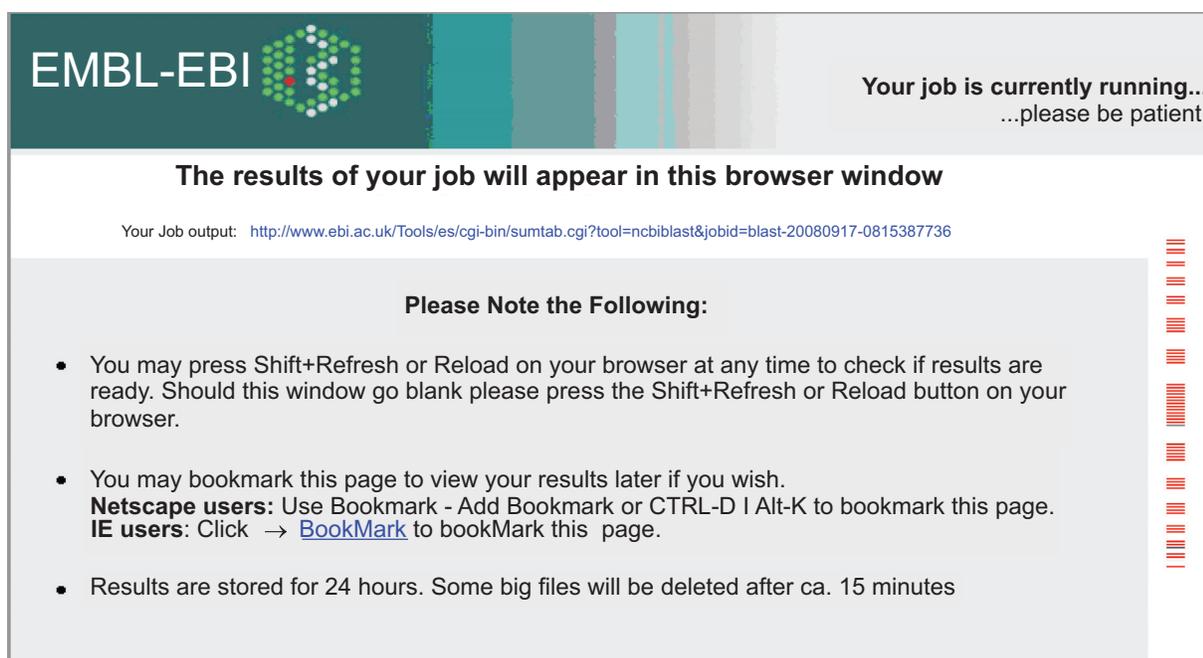


Figure 1. Page refresh in EBI-NCBI BLAST.

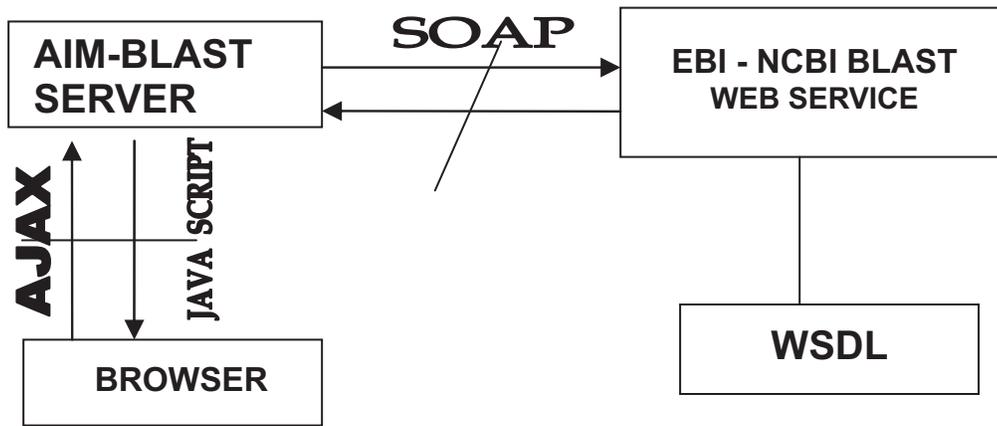


Figure 2. Overall process chart of AIM-BLAST.

In this AJAX pattern (Fig. 3), the XMLHttpRequest object binds to a callback Javascript function and then sends a POST or GET request to the server asynchronously. The handler function monitors the ready State property of XMLHttpRequest that changes as the request goes through and the response is received. Until the ready State becomes 4 (meaning that the response has been completely received) a progress bar is displayed to signal the progress of the long running process. Once ready State is 4, the callback handler gleans the results out of the response XML and displays the result by DOM manipulation without page refresh. Thus, in AIM-BLAST, the unpleasing page refresh menace that is very common in any other blast tools, is greatly controlled resulting

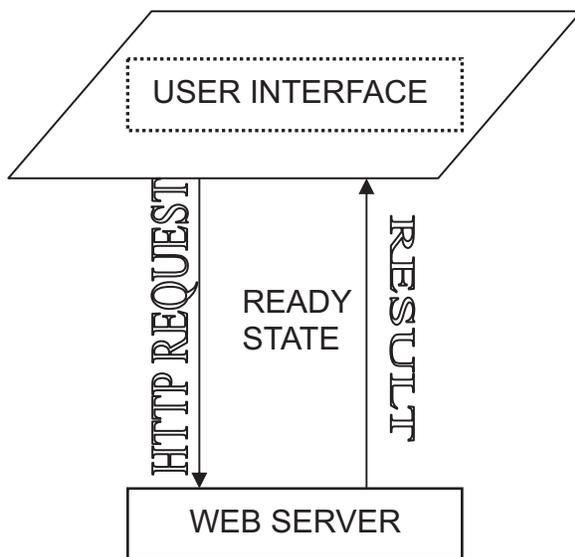


Figure 3. AJAX technology in AIM-BLAST.

in the minimized bandwidth consumption but, still performs effectively.

This tool makes it possible to perform the annotation of an entire genome at a single submit. The input for this tool is the protein sequences in FASTA format. As soon as the sequences are submitted to AIM-BLAST, the sequences are forwarded to the EBI server where each sequence is individually compared against all the genome databases. When the analysis is being carried out, a simple progress bar appears on the screen without refreshing the entire page. The Perl server of AIM-BLAST utilizes the SOAP⁷ web services of EMBL-EBI, (European Molecular Biology Laboratory- European Bioinformatics Institute) to fetch the results from the Blast server.

Once the results are ready, AIM-BLAST will carryout the automatic parsing using some special filtering process that can expertly handle the baggy Blast results of the sequences and produce one hit for one sequence. The filtering process is performed in two parts. The first part of filtering is carried out to choose the Blast hits that satisfy the values of all parameters including Blast score, the length and orientation of the hits, the percentage identity, percentage similarity and E-values. The second part of the process involves the further cleaning of the functions with any negative terms, functions that do not have any clear scientific evidence, such as predicted, putative, probable, hypothetical, conserved hypothetical and unknown. Thus, this filtering process of results in AIM-BLAST remains a powerful means of reducing the possibility of errors while choosing single significant function from the massive

Blast hits. The Results of AIM-BLAST appear in a simple and easily interpretable table format. In case the user is not satisfied with the AIM-BLAST result and required to manually interpret the bulky blast hits for each sequence then this is also possible with AIM-BLAST. The user can click the result option of every sequence that is available in the AIM-BLAST result table and a new window appears in the browser that shows the entire blast result of the clicked sequence so as to facilitate manual interpretation. Additionally, there is also an option within this tool to save the results in PDF format for any further analysis. With all these features, AIM-BLAST remains a user-friendly and an efficient tool for performing sequence similarity search to multiple sequences.

Results and Discussion

To evaluate the efficiency of AIM-BLAST, we have compared the performance of this tool with the regular NCBI BLAST service offered by the European Bioinformatics Institute using the real time date. A sample set 30 protein sequences of varying length from *E.coli K12 strain* are simultaneously analyzed using the AIM-BLAST and the regular EBI-NCBI Blast.

Minimized Band width Consumption

Both the AIM-BLAST and the EBI-NCBI BLAST are run in the Firefox Web browser and the HttpFox (<https://addons.mozilla.org/en-US/firefox/addon/6647>), a Firefox add-on is operated at the backend to measure the amount of bandwidth consumption during the analysis. As soon as the analysis of the entire set of sequences is completed, the loads of bytes sent and received for the sample protein

sequences is tabulated (Table 1) for comparison. As per the resultant Table, it is observed that EBI-NCBI Blast consumed an overall data transfer of 12.62 Mega Bytes viz. 0.8 MB of data sent to the server and 11.80 MB of data received from the server for analyzing just 30 sample sequences. Whereas, AIM-BLAST consumed only 0.08 Mega Bytes of data transfer viz. 0.049 MB of data sent to the server and 0.031 MB of data received from the server. Moreover, as in regular blast service, the extensive book keeping at the server to keep track of jobs and job-ids is not required for AIM-BLAST and this ensures that this tool reduces the superfluous network traffic and saves bandwidth. Additionally, In AIM-BLAST, the precarious and visually unappealing page refresh, which is common in any regular BLAST service, has been replaced by a simple progress bar that remains an effective user interface paradigm.

Saves Man-Power and Man-Hour

Further, in AIM-BLAST, as soon as the analysis is completed, the results are directly available in a table that can be saved as a PDF file in no time. The results displayed by AIM-BLAST are clean functions that are automatically filtered from a huge number of Blast hits thus saving the hectic human parsing and the enormous handling time for choosing one appropriate hit for every sequence. Hence the overall process time for the sample set of sequences is only 12 Minutes and 03 seconds. Whereas, in EBI-NCBI BLAST, as soon as the Blast results for each sequence are available, they are manually interpreted, one appropriate hit is chosen based on various parameters and then the selected function is copied and pasted to a local file for further analysis. This makes the process much frenzied and took an

Table 1. Comparison of total data transfer and the overall processing time between EBI-NCBI BLAST and AIM-BLAST for the sample set of sequences.

Tools	EBI-NCBI BLAST	AIM-BLAST
Data transfer during analysis		
Data Sent [in Bytes]	818769	49488
Data Received [in Bytes]	11800797	30596
Total Data Transfer [in Bytes]	12619566	80084
Total Data Transfer [in Mega Bytes]	12. 619566	0.080084
Processing time during analysis		
Time [in Minutes]	96.32	12.03

overall course time of 96 minutes and 32 seconds for only 30 sequences. Therefore AIM-BLAST remains a simple, but novel, tool to carry out sequence similarity searches of multiple biological sequences more quickly than any other blast services besides consuming limited bandwidth.

Conclusion

We present AIM-BLAST as one of the most appropriate and coordinated programs in which we have overcome the challenge to analyze the overwhelming biological sequences faster than it is carried out using other blast services. Above all, AIM-BLAST produces the results in a pleasing and presentable manner that simply provides a better user experience. Henceforth, AIM-BLAST will find a vital role in the Next Generation genomic researches.

Authors Contributions

Conceived and designed the concept: GR, GA. Programming: RS, GA. Testing the tool: GR, KS, GA.

Tool Evaluation: AD, KK. Data Analysis: GA, RS. Wrote the Paper: KS, GA. Revised the Paper: GR, AD, KK.

Disclosure

The authors report no conflicts of interest.

References

1. Kim Carter, Matthew Bellgard. MASV—Multiple (BLAST) Annotation System Viewer. *Bioinformatics*. 2003;19(17):2313–2315.
2. Sivashankari S, Shanmughavel P. Functional annotation of hypothetical proteins—A review *Bioinformation*. 2006;29:1(8):335–8.
3. Diana M. Downs Genomics and Bacterial Metabolism. *Curr Issues Mol Biol*. 2003;5(1):17–25.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–410.
5. Leslie Klis McNeil, Claudia Reich, Ramy K. Aziz, et al. The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation. *Nucleic Acids Research*. 2006;D347–D353.
6. Garrett J. AJAX: A new approach to web applications. Adaptive path, <http://www.adaptivepath.com/publications/essays/archives/000385.php>. 2005.
7. Pillai S, Silventoinen V, Kallio K, et al. SOAP-based services provided by the European Bioinformatics Institute. *Nucleic Acids Research*. 2005;33:W25–W28.