

easySEARCH: A user-friendly bioinformatics program that enables BLAST searching with a massive number of query sequences

Dong-Wook Kim^{1†}, Ryong Nam Kim^{1†}, Dae-Soo Kim^{1†}, Sang-Haeng Choi¹, Sung-Hwa Chae³, Hong-Seog Park^{1, 2*}

¹Genome Resource Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-806, Korea; ²University of Science and Technology (UST), Daejeon 305-333, Korea; ³Research Institute of GnC Bio Co. Ltd, Daejeon 305-150, South Korea; Hong-Seog Park – Email: hspark@kribb.re.kr; *Corresponding author

†- These authors contributed equally to this work

Received August 06, 2012; Accepted August 15, 2012; Published August 24, 2012

Abstract:

Many biologists are familiar with BLAST (Basic Local Alignment Search Tool). A major difficulty with BLAST is searching a massive number of queries without depending on professional help with scripting languages. Hence, we describe the development of an interface for BLAST to perform sequence search for a set of query sequences at one instance. This software interface runs on all Windows-based personal computers (PCs) and can be widely used by biologists who are not familiar with professional informatics command languages and yet want to perform massive sequence analysis. easySEARCH is freely available as a stand-alone program.

Availability: http://210.219.44.213/easysearch_down.html.

Key words: easySEARCH, Bioinformatics program

Background:

BLAST (Basic Local Alignment Search Tool) has been utilized in nearly every branch of biology, far beyond the scope of molecular genetics, molecular biology and protein biochemistry, and this tool has made great contributions to many scientific fields since its development [1, 2]. Currently, the work of most biologists, bioinformaticians, evolutionists and medical scientists cannot progress without the use of BLAST.

Although the use of BLAST is common in many scientific disciplines, bioinformaticians and biologists use BLAST in very different ways. Bioinformaticians can perform a one-round BLAST search with a massive number of query sequences using an informatics command language. In contrast, biologists are

often not familiar with informatics command languages and usually perform a one-round NCBI BLAST search using one query or a few (≤ 20) queries. If biologists want to analyze a massive number of query sequences, they must carry out thousands of the one-round BLAST search with one query, which can be very time consuming and tedious. Thus, there is a need for software that can run a one-round BLAST search with a massive number of queries and also can be used by biologists who are not proficient in informatics program languages.

Here, we present “easySEARCH,” a newly developed bioinformatics program that performs one-round BLAST search with a massive number of queries using a mouse-click style interface. Due to its user-friendliness and convenience,

BIOINFORMATION

open access

easySEARCH can be widely used by biologists, bioinformaticians, medical scientists and evolutionists.

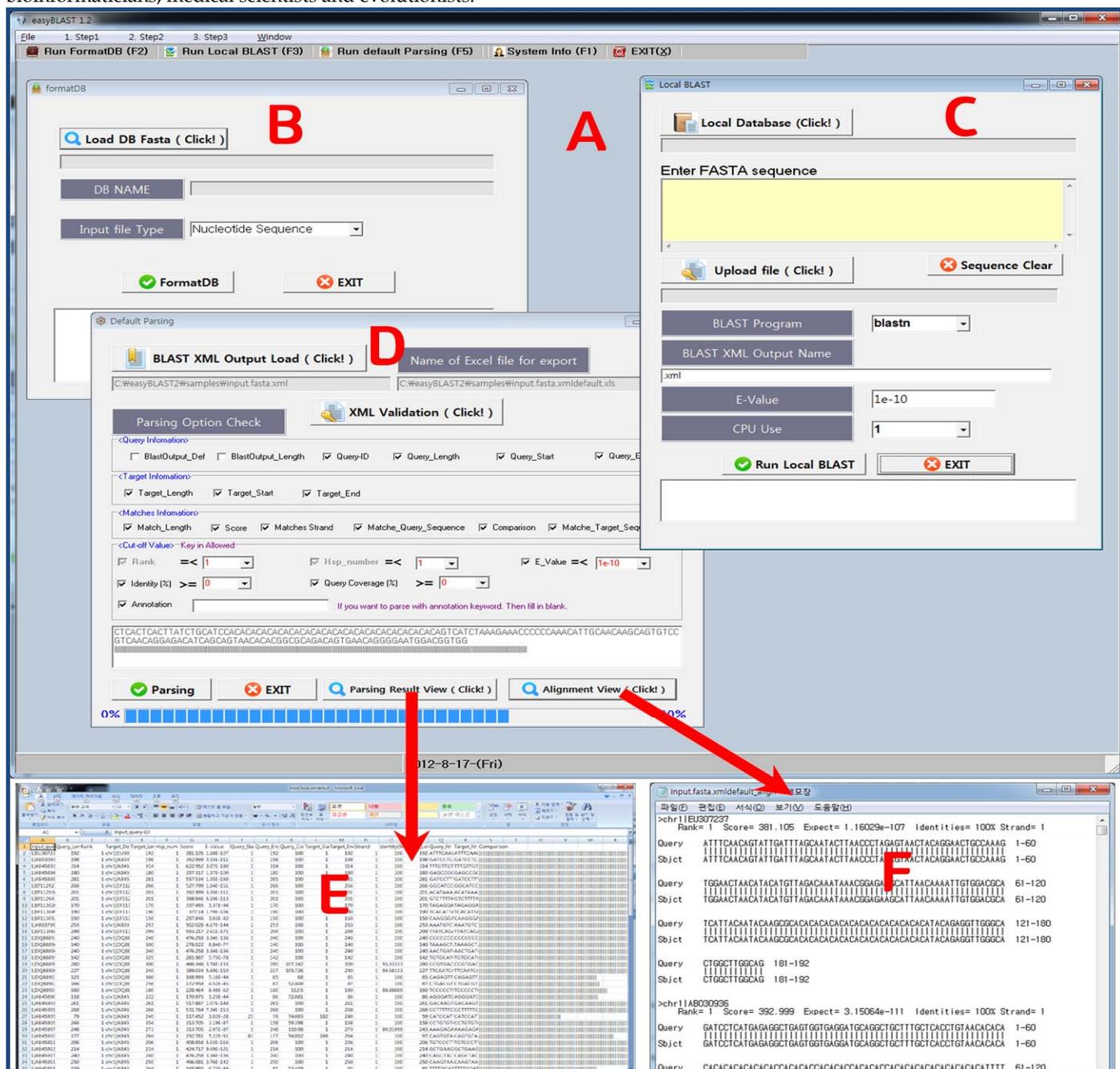


Figure 1: BLAST Process of the easyBLAST. The windows A, B, C and D are the windows that have been used during the BLAST process by the easyBLAST program. The windows E and F are the windows showing the results produced during the BLAST process. The detailed explanations of those windows A, B, C, D, E and F are described in the text of this paper.

Methodology:

This software can be run on personal computers (PCs) that run Windows. If the user has installed this software on his/her PC from the above-mentioned website, the user will find a directory called "easySEARCH" on the C drive; the easySEARCH software file (red icon) can be found in this directory.

The method for using this software is very user-friendly and convenient. If the easySEARCH software file (red icon) is opened, window A will appear (Figure 1). First, either the

"FormatDB F2" button (in the pull-down menu of Step 1 in window A) or the "Run FormatDB (F2)" button (in window A) can be pressed to format a database. In the open window, window B (Figure 1), the user can press the "Load DB Fasta (Click!)" button to load a sequence database (for example, the "db" database in the "samples" file provided in the easySEARCH software package) into easySEARCH. The user can then choose "Nucleotide Sequence" or "Protein Sequence" from the pull-down menu of the "Input file type" depending on the user's sequence database. If the "FormatDB" button is pressed, the database will be formatted. After finishing this

process, the message shows "Complete FormatDB. Go to BLAST Menu!!" will appear, and the "OK" button should be pressed.

Next, the "Local BLAST F3" button (in the pull-down menu of Step2) or the "Run Local BLAST (F3)" button should be pressed. In the open window, window C (**Figure 1**), the button "Upload file (Click!)" will load a query sequence file (for example, the "input" query data in the "samples" file provided in the easySEARCH software package) into easySEARCH. Otherwise, the user can directly paste query sequences into the yellow box just below the command sentence "Enter FASTA sequence" in window C. The user can then check whether the appropriate BLAST program (blastn, blastp, blastx, tblastn, tblastp and tblastx), E-value (1e-10 as a default value) and CPU number (1 in the case that one PC is used) have been chosen. The "Run Local BLAST" button will initiate the BLAST run. After completion, the message shows "Complete BLAST". Go to Parsing menu!!" will appear, and the "OK" button should be pressed. If the user needs to use NCBI BLAST, the user can press the "NCBI BLAST F4" button instead of the "Local BLAST F3" button in the pull-down menu of Step2. Subsequently, a window will appear in which an NCBI BLAST (based on the NCBI web-server) can be performed.

Finally, the "Default Parsing F5" button (in the pull-down menu of Step3) or the "Run default Parsing (F5)" button should be pressed. In the open window, window D (**Figure 1**), the appropriate parsing options, which will affect the contents of the BLAST result files, can be selected. Specifically, if certain contents of interest need to be extracted from a whole BLAST result file, a keyword (representing the contents of interest) can be inserted into the "Annotation" box as an optional item. After parsing (by pressing the "Parsing" button), the "OK" button should be pressed when the user is prompted with the message, "Complete Parsing". Next, a parsing result (window E) and an alignment file (window F) can be obtained by pressing the "Parsing Result View (click!)" and "Alignment View (Click!)" buttons, respectively (**Figure 1**). If more professional options are needed for parsing, the "High Level Parsing" window can be

opened by pressing the button "High Level Parsing F6" in the pull-down menu of Step 3.

Input:

A query sequence file and a sequence database file must be available in FASTA format. It should be noted that the number of query sequences for the one-round BLAST process is not limited in the easySEARCH interface. The NCBI NR database could be downloaded from http://210.219.44.213/easyblast_down.html as and when required.

Output:

The outputs of easySEARCH include the parsing result (Excel format) and the alignment file (Notepad format) (**Figure 1E & F**).

Feature:

easySEARCH has been written using the Delphi programming language (version 5.0). This software can enhance the speed of the BLAST process by parsing line by line with XML validation.

Conclusion:

We describe easySEARCH to implement BLAST searches with a massive set of query sequences. The use of easySEARCH bypasses the need for command languages because it utilizes a mouse-click style interface. The practical use of the user-friendly easySEARCH finds utility by a number of end users.

Acknowledgement:

We thank the members of the Genome Resource Center (GRC) at the Korea Research Institute of Bioscience and Biotechnology (KRIBB) for their active assistance in conducting this research project. *Funding:* This work was supported by the Ministry of Education, Science and Technology (MEST) [grant 2009-0084206] and KRIBB [grant KGM4131211].

References:

- [1] Altschul SF *et al.* *Nucleic Acids Res.* 1997 **25**: 3389 [PMID: 9254694]
- [2] Altschul SF, *J Mol Biol.* 1991 **219**: 555 [PMID: 2051488]

Edited by P Kanguane

Citation: Wook Kim *et al.* *Bioinformatics* 8(16): 792-794 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited