Smart eLab - volume 8, anno 2016

Highlighter

# Visual Blast.

Giovanni Mele*[a]

BLAST (Basic Local Alignment Search Tool) is a popular program that retrieves a library of sequences that resemble the query. The major problem experienced by new users of BLAST revolves around constructing syntactically and semantically correct command line, getting input files into acceptable formats and assessing the output. Here, we present Visual BLAST a Graphical User Interface to perform BLAST searches. Visual BLAST aims to make BLAST searches accessible to a wider audience with no bioinformatics skill and to facilitate usage among the existing.

**Keywords**: Sequence Analysis, Graphic User Interface, Software.

## 1 Introduction

BLAST remains one of the most widely tools used in computational biology. This popular common line program was developed by Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, and David J. Lipman[1]. BLAST allows comparing a DNA or protein sequence query with a database of sequences, and consents retrieving a library of sequences that resemble the query. BLAST addresses fundamental problems needed in computational biology research. In fact, BLAST can be used for several purposes. These include identifying species, locating domains, establishing phylogeny, DNA mapping, and comparisons.

The BLAST program can either be free available for download at http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download and run as a command-line utility "blastall" or accessed for over the web at http://blast.ncbi.nlm.nih.gov/Blast.cgi. The BLAST web server, hosted by the NCBI, allows anyone with a web browser to perform similarity searches against constantly updated databases of proteins and DNA that include most of the newly sequenced organisms.

In the genomic era, the necessity to search several thousand of sequences in a single query makes the utilization of web-hosted programs not feasible for both low speed over internet and for problems correlated with server traffic and stability. At the same time, the BLAST utility that run on local computers has the main drawback in the complex and long command-line string required to run the analyses. In fact, command-line computing environments are very challenging for users without programming experience. Consequently, there is a pressing need for a menu-driven or Graphical User Interface to allow biologists to access the methodology without becoming programmers. To address these and others issues, Visual BLAST, a Graphical User Interface for BLAST searches was developed.

## 2 Description

Visual BLAST is a user-friendly GUI written in C# and it was developed to simplify the parameters setting for BLAST searches. Visual BLAST implements BLAST allowing the utilization of blank space in the input and output file name and in the folder path. Moreover, Visual BLAST places in the first line of the output tabular text file the headers for each output option selected by the user. Finally, the GUI of Visual BLAST save the last user input settings allowing easier and more efficient searches when running multiple analyses with the same parameters. The application consists of one window with two tabs: *Blast Search and Database Management* (Figure 1).

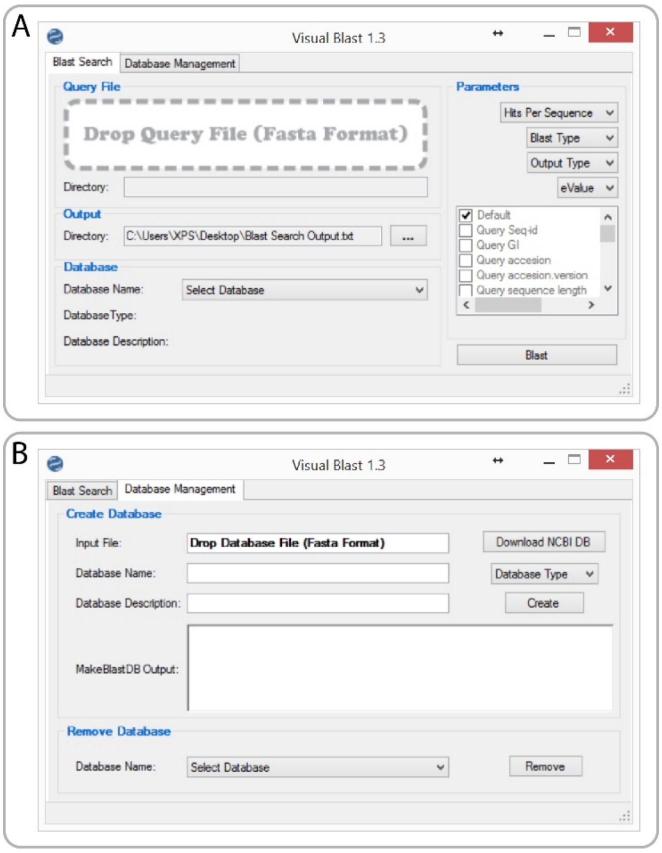
### 2.1 Blast Tab

The *Blast Search* tab consists of four panels (Fig. 1A). The Query File panel contains a drag and drop box that accepts the input file for the search in FASTA format. On the right, the *Parameters* panel consists of four dropdown boxes and a checklist box. The dropdown boxes allow setting the number of hits for each input sequence, the type of search (BLASTn for nucleotide query on nucleotide database, BLASTp for aminoacid query on protein database and BLASTx for nucleotide query on protein database), the output file (Table 1) and the minimum e-Value cutoff. In case that the user selects as output the file format option 6, 7 or 10, the checklist box allows to select a custom format specifiers (Table 2). The default specifiers are: *qseqid, sseqid, pident length mismatch gapopen qstart qend sstart send*

---

[a] *Institute of Agricultural Biology and Biotechnology, National Council of Research, Via Salaria Km. 29.300 00015 Monterotondo Scalo, Rome, Italy.*

* **Contact:** melegio@ibba.cnr.it.



Directory: C:\Users\XPS\Desktop\Blast Search Output.txt

☑ Default
☐ Query Seq-id
☐ Query GI
☐ Query accesion
☐ Query accesion.version
☐ Query sequence length

**Fig. 1** Visual BLAST Tabs. A) Blast Search Tab. B) Database Management Tab.

*evalue*. The Output panel allows selecting the output directory and output file name. The *Database* panel allows the selection of the database for the query.

**Tabella 1** Output file format.

| Selection | Output file format description |
|---|---|
| 0 | Pairwise |
| 1 | Query-anchored showing identities |
| 2 | Query-anchored no identities |
| 3 | Flat query-anchored, show identities |
| 4 | Flat query-anchored, no identities |
| 5 | XML Blast output |
| 6 | Tabular with headers |
| 7 | Tabular with comment lines |
| 8 | Text ASN.1 |
| 9 | Binary ASN.1 |
| 10 | Comma-separated values |
| 11 | BLAST archive format (ASN.1) |

### 2.2 Database Tab

The *Database Management* tab allows the databases administration and consists of an upper panel for the creation of a new database and a lower part that allows the removal of unwanted databases (Figure. 1B). The database creation section comprises of a drag and drop box for the creation of the new database, of two text boxes for the database information and of a drop down box for the selection of database type (nucleotide or proteic). Moreover, the *Download NCBI DB* button open the NCBI FTP page for the downloading of the NCBI curated prebuild databases. The lower panel of the tab allows the selection and the deletion of unwanted database.

**Tabella 2** Custom format specifiers.

| Specifiers | Description |
|---|---|
| qseqid | Query Seq-id |
| qgi | Query GI |
| qacc | Query accesion |
| qaccver | Query accesion.version |
| sseqid | Subject Seq-id |
| sallseqid | All subject Seq-id(s), separated by a ';' |
| sgi | Subject GI |
| sallgi | All subject GIs |
| sacc | Subject accession |
| saccver | Subject accession.version |
| sallacc | All subject accessions |
| qstart | Start of alignment in query |
| qend | End of alignment in query |
| sstart | Start of alignment in subject |
| send | End of alignment in subject |
| qseq | Aligned part of query sequence |
| sseq | Aligned part of subject sequence |
| evalue | Expect value |
| bitscore | Bit score |
| score | Raw score |
| length | Alignment length |
| pident | Percentage of identical matches |
| nident | Number of identical matches |
| mismatch | Number of mismatches |
| positive | Number of positive-scoring matches |
| gapopen | Number of gap openings |
| gaps | Total number of gaps |
| ppos | Percentage of positive-scoring matches |
| frames | Query and subject frames separated by a '/' |
| qframe | Query frame |
| sframe | Subject frame |
| btop | Blast traceback operations (BTOP) |
| qseqid | Query Seq-id |

## 3 Conclusion

Visual BLAST executable is free of charge software that runs on 32 and 64 bits Windows platform and it is tune up for Windows XP, 7 and 8. It is distributed in pre-packaged self-extracting installer for installation on local computers. Visual BLAST is a user friendly and flexible GUI software to retrieve a library of sequences that resemble the query. The easy to use GUI Interface benefits a wide audience for a fast and effective analysis.

**Availability and Implementation:** Visual BLAST executable is freely available on the web page of the Institute of Agricultural Biology and Biotechnology of National Council of Research (http://www.ibba.mlib.cnr.it/Visual_Blast.html), Softpedia (http://www.softpedia.com/get/Science-CAD/Visual-Blast.shtml) and Softtonic (http://visual-blast.en.softonic.com). This software is design to be fully compatible with Windows XP, 7, 8 and 10 environments.

## 4 Acknowledgements

Special thanks to Donato Giannino e Giulio Testone for the critical reading of the paper. This work was supported by a dedicated grant from the Italian Ministry of Economy and Finance to the National Research Council for the project "Innovazione e Sviluppo del Mezzogiorno - Conoscenze Integrate per Sostenibilità ed Innovazione del Made in Italy Agroalimentare - Legge n. 191/2009"

## Riferimenti bibliografici

1 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool, Journal of Molecular Biology 215 (3) (1990) 403 – 410. `doi:10.1016/S0022-2836(05)80360-2`.