

Improving your BLAST+ experience with CrocoBLAST

Ravi Jose Tristao Ramos¹

1 CEITEC - CENTRAL EUROPEAN INSTITUTE OF TECHNOLOGY

Abstract

CrocoBLAST is a tool designed for improving BLAST+ speed and user experience, bringing it closer to the end users. CrocoBLAST was benchmarked on 4 different computers (including a low-end desktop and a modern server), with 11 different datasets (including a proteome, a metagenome and diverse NGS read sets) comprising the 5 main BLAST+ applications (blastn, blastp, blastx, tblastn, tblastx). The benchmark shows that large alignments that would require a dedicated computer for several weeks with NCBI BLAST+, as is commonly the case for NGS data, can be run overnight with CrocoBLAST. Additionally, CrocoBLAST provides enhanced user experience features such as: real-time information regarding calculation progress and remaining run time; access to partial alignment results; queuing, pausing, and resuming BLAST+ calculations without information loss. CrocoBLAST was implemented as a friendly software layer between the end user and BLAST+, and is compatible with any BLAST+ version, providing identical results to those of BLAST+. Furthermore, CrocoBLAST allows for using files in FASTQ format as query for the alignment; easily downloading pre-formatted BLAST+ databases from NCBI; adding existing pre-formatted BLAST+ databases; or easily creating new databases directly from FASTA or FASTQ files. All functionality is available both by command line and by a graphical user interface developed in Java. CrocoBLAST is freely available for download, and comes with a user manual exemplifying the usage of all its commands, including image examples of the graphical user interface (webchem.ncbr.muni.cz/Platform/App/CrocoBLAST). No installation or user registration is required. CrocoBLAST is available for Linux and Windows; Mac OS X users can run CrocoBLAST with good performance within a Linux virtual machine.

Funding: CEITEC - Central European Institute of Technology