

BIOINFORMATICS

Electronic edition <http://bioinformatics.oupjournals.org>

VOLUME 19
NUMBER 17
NOVEMBER 22 2003
PAGES 2313–2315

DOI: 10.1093/bioinformatics/btg301

MASV—Multiple (BLAST) Annotation System Viewer

Kim Carter and Matthew Bellgard*

Centre for Bioinformatics and Biological Computing, School of Information Technology, Murdoch University, Murdoch 6150, WA, Australia

Received on November 27, 2002; revised on March 10, 2003; accepted on May 7, 2003

*To whom correspondence should be addressed.



GO BACK

CLOSE FILE

Abstract

Summary: *Multiple (BLAST) Annotation System Viewer (MASV) is a tool designed to aid in the annotation of genomic sequences. MASV enables the researcher to compare and analyse differences in annotation and analysis, resulting from changes in databases, analysis program parameters and results. This provides a unique capability for the user to conduct further bioinformatics analysis from the information obtained.*

Availability: <http://cbbc.murdoch.edu.au/projects/masv/>

Contact: kCarter@cbbc.murdoch.edu.au

Abstract

Introduction

Program Overview

References



GO BACK

CLOSE FILE

Introduction

The last 10 years have seen a tremendous increase in the volume of biological data being stored and manipulated worldwide. Along with the increase in volume of data, there has been an increase in data related errors (Krawetz, 1989; Bhatia *et al.*, 1997; Carter *et al.*, 2001). With such large amounts of data available, it is becoming essential to have the ability to examine and compare results in a graphical form, rather than text. A graphical presentation of results can show the location of important information in a far less cluttered fashion, enabling rapid analysis and interpretation of results. A system that also recorded the details of the analysis at every step could also minimize human error resulting from having to analyse large amounts of textual data, as well as enable visual comparison of results at a later date.

Sequences and annotations contained in sequence repositories, such as GenBank (Benson *et al.*, 2002), are being continually updated. Related systems such as DAS (Stein *et al.*, 2002, <http://www.biodas.org>) can allow a user to examine annotations from different sources; however, the current graphical clients are still under development and rely on remote DAS servers to be running. NIX (<http://www.hgmp.mrc.ac.uk/NIX/>) is a web based tool that allows registered users to view the results of running multiple DNA analysis on a given sequence; however, the system restricts users from selecting different program parameters (defaults are already assigned) and is restricted in sequence size. In addition to using

Abstract

Introduction

Program Overview

References



GO BACK

CLOSE FILE

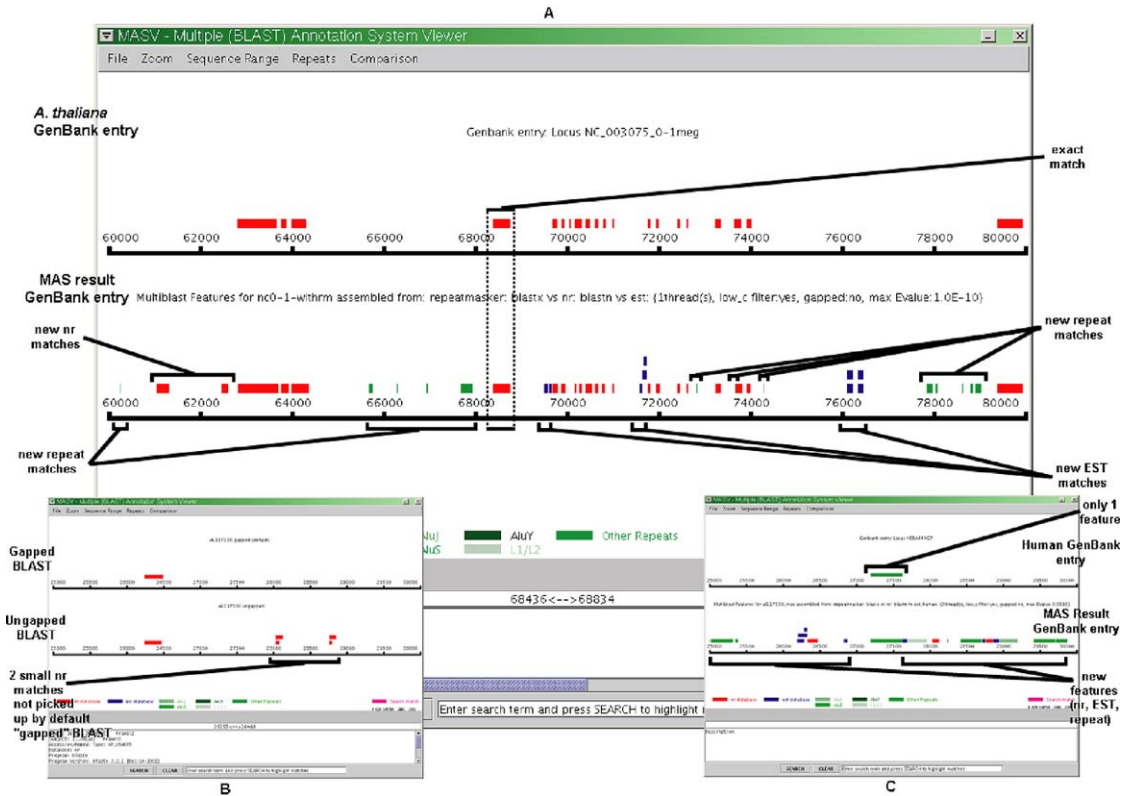


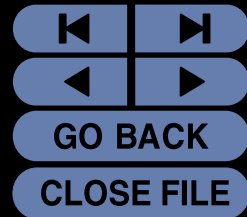
Fig. 1. MASV Screenshots. **(A)** *A.thaliana* GenBank annotation versus MAS analysis. **(B)** Different BLAST parameters. **(C)** Human GenBank annotation versus MAS analysis.

Abstract

Introduction

Program Overview

References



web-based bioinformatics tools, researchers desire to conduct and control their own analysis and verify and review results locally (Bellgard *et al.*, 1999). It is essential for a researcher to have the ability to compare and analyse third party created results with those conducted locally. Carter *et al.* (2001) describes the importance of keeping up-to-date information (sequences and annotations) and introduces the Multiple (BLAST) Annotation System (MAS) for automated systematic feature identification, vital for re-annotation of genomic sequences. We have designed a system, referred to as MASV, which enables comparison and analysis of locally and remotely created sequence analysis and annotations, as we have recognized that this type of tool is not currently available. MASV facilitates analysis of multiple results simultaneously, in a visual form to enable further analysis to be conducted.

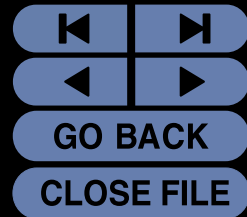
Figure 1 shows the results of the MASV system in three different situations. Figure 1A shows a region of a comparison between an *Arabidopsis thaliana* GenBank accession annotation (NC_003075.2, submitted August 2002) and the result of running the same sequence through MAS (conducted Nov. 2002). As the results show, numerous features [we refer to a 'feature' as a partial fragment or complete match with a repetitive, nr (protein) or expressed sequence tag (EST) sequence] have been identified with MAS analyses that do not appear in the existing GenBank entry annotation. Figure 1B shows a comparison of the analysis results of a particular GenBank accession, where the top sequence is

Abstract

Introduction

Program Overview

References



a gapped (default) BLAST (Altschul *et al.*, 1990) and the bottom is an ungapped BLAST. The bottom (ungapped BLAST) sequence result has two small exons of the same gene that were not found using the default BLAST. Figure 1C shows a region of a comparison of a human GenBank accession annotation versus the MAS analysis of the same sequence, which displays numerous new features being identified. For each MAS result, the input sequence was examined with RepeatMasker (Smit and Green, 2002 <http://ftp.genome.washington.edu/RM/RepeatMasker.html> – Version 2002/07/13, RepBase 7.7), a BLASTx (Version 2.2.4) versus nr protein database (1 220 597 sequences) and a BLASTn (Version 2.2.4) versus EST (human, 4 844 460 sequences). As these examples illustrate, comparing analysis and annotations graphically allows them to be examined and refined simply and efficiently.

Abstract

Introduction

Program Overview

References



GO BACK

CLOSE FILE

Program Overview

MASV is a graphical annotation viewer allowing single viewing or side by side comparison of results (shown in [Fig. 1](#)), currently up to four. The user can select a level of zoom and sequence range to view different sections of the annotation contained in the input file(s). Features are shown in different colours, and when selected, the underlying match (if available) is shown. When two annotation files are input to MASV, a simple comparison is made to determine if features, given a start and end position, in the first sequence correspond with those in the second sequence, and vice versa. Comparisons of features can result in exact matches, slight positional mismatches (with same feature) or different features at (about) the same sequence position, which are shown graphically (not shown in [Fig. 1](#)).

MASV is written in Java (as a Java Applet) and is accessible via a web browser (requires Java 1.4). Currently MASV accepts BLAST-XML results, GenBank and GenBank (sequence) XML files, and MultiBLAST XML ([Carter et al., 2001](#)) files. At present the viewer displays protein features, EST and repeat sequences, although additional feature displays can be added.

[Abstract](#)

[Introduction](#)

[Program Overview](#)

[References](#)



GO BACK

CLOSE FILE

References

- Altschul,S.F., Gish,W. and Miller,W. (1990) Basic Local Alignment Search Tool. *J. Mol. Biol.*, **215**, 403–410. [MEDLINE Abstract](#)
- Bellgard,M.I., Hiew,H.L., Hunter,A. and Wiebrands,M. (1999) ORBIT: an integrated environment for user customised bioinformatics tools. *Bioinformatics*, **15**, 847–851. [MEDLINE Abstract](#)
- Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J. *et al.* (2002) GenBank. *Nucleic Acids Res.*, **30**, 17–20. [MEDLINE Abstract](#)
- Bhatia,U., Robison,K., Gilbert,W., Klenk,H., White,O. and Venter,J.C. (1997) Dealing with database explosion: a cautionary note. *Science*, **276**, 1724–1725. [MEDLINE Abstract](#)
- Carter,K., Oka,A., Gen,T. and Bellgard,M.I. (2001) Bioinformatics issues for automating the annotation of genomic sequences. *Genome Inform.*, **12**, 204–211.
- Krawetz,S.A. (1989) Sequence errors described in GenBank: a means to determine the accuracy of DNA science interpretation. *Nucleic Acids Res.*, **17**, 3951–3957 [MEDLINE Abstract](#)
- Pearson,W.R. and Lipman,D.J. (1988) Fasta. *PNAS*, **85**, 2444–2448. [MEDLINE Abstract](#)
- Smit,A.F.A. and Green,P. (2002) RepeatMasker [online].
- Stein,L.D., Eddy,S. and Dowell,R. (2002) Distributed Sequence Annotation System (DAS) [online].

[Abstract](#)

[Introduction](#)

[Program Overview](#)

[References](#)



GO BACK

CLOSE FILE