Bi◯techvana

Biotechvana Bioinformatics

TIME: a sequence editor for the molecular analysis of large DNA and protein sequence samples

Muñoz-Pomer, A.[1,2], Futami, R.[1], Covelli, L.[1], Domínguez-Escribà, L.[1], Bernet, G.P.[1], Sempere, J.M.[2], Moya, A.[3,4], Llorens, C.[1]

1. Biotechvana, Parc Cientific de la Universitat de València
2. Departamento de Sistemas Informáticos y Computación (DSIC), Universitat Po-litècnica de València
3. Unidad Mixta de Investigación en Genómica y Salud del Centro Superior de In-vestigación en Salud Pública (CSISP)-Universitat de València (Instituto Cavanilles de Biodiversidad y Biología Evolutiva)
4. CIBER en Epidemiologia y Salud Pública (CIBEResp)

Corresponding author: carlos.llorens@biotechvana.com

Availability: Available online February 1, 2011 at http:// www.biotechvana.com/software/time

Background: In this article we introduce the release of TIME (Tool for In-place Molecular Editing), a sequence editor devoted to the analysis of large nucleotide and protein sequences such as chromosomes, genomic contigs and their encoded protein products.

Remarks: TIME offers a variety of functions for editing, translating and managing single and multiple sequence files. One of TIME´s main features is its ability to process large sequences up to 25 megabases. It includes search capabilities for retrieving open reading frames (ORFs) and their coordinates as well as other user-defined motifs such as restriction, binding, priming sites, etc.

Availability: TIME is distributed by Biotech Vana SL at the following [URL 1]. A 15 days trial version is available.

Keywords: Bioinformatics | Computational biology

INTRODUCTION

A key element in the rapid advance of molecular biology and omics research has been the development of algorithms and computational methods for processing biological data, accompanied by an exponential growth in computing power value (MIPS per dollar). As sequencing throughput continues to grow owing to the advent of next generation DNA sequencing methods [1-6], computational tools for handling large volumes of biological data become necessary. One of the most common troubles met during genomic sequence analyses is that, owing to the great size of a single chromosome, research is carried out in a fragmented set of sequences [7]. In this paper we present TIME, a software that tackles this problem through efficient memory management when working with sequences that cannot fit into the main memory.

OVERVIEW

Features: TIME is a powerful and versatile tool that allows in-place editing of both nucleotide and amino acid sequences up to 25×10^6 characters, enough even for cull chromosomes.

TIME software has been programmed to provide easy-to-use features for both basic and advanced analyses. The functions are organized in a hybrid menu-tool bar and have been designed into a common interface that enables easy and logical handling, retrieval, storage and results display.

TIME accepts both single and multiple sequence files in FASTA format, and allows sequences to be imported straightforwardly from databases such as GenBank [9] or EMBL [10] through copy-pasting into a blank TIME data-sheet. Users have also the option to unlock and edit any sequence either by typing or with the aforementioned cut, copy and paste commands. Sequence geometry and orientation can also be managed. As shown in Figure 1, nucleotide sequences can be translated to all six reading frames, with the start and stop codons highlighted in user-customizable colors. The genetic code for translation can be easily modified and saved into a simple plain text format, so it may be recovered in later sessions.
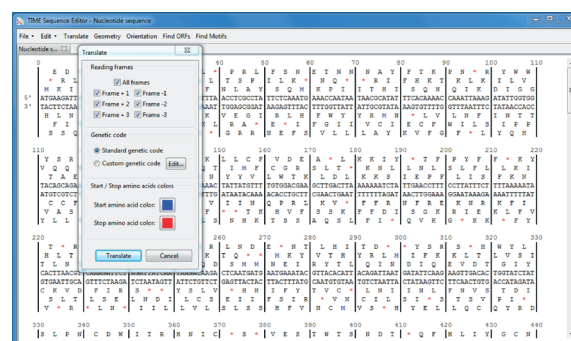

Figure 1. TIME screenshot showing the six frame translation of a nucleotide sequence

TIME includes flexible tools for finding ORFs and motifs checking both orientations of the translated sequence. ORFs can be required to have a minimum length and the start and stop codons are specified. As shown in Figure 2, motifs may be searched either in single occurrences or in clusters. In the latter case, parameters such as cluster size, minimum number of motifs within a cluster and/or overlapping clusters can all be specified from the motif editor. ORFs and motifs can be saved and exported to CSV spreadsheet files or as a FASTA file.
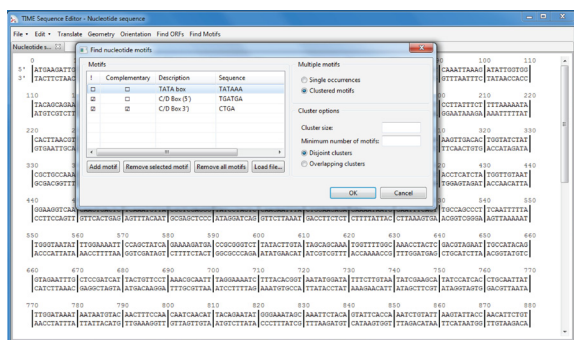
Figure 2. TIME screenshot showing the multiple motif editor.

its capacity to process sequences up to a several megabases in size. TIME is distributed both as a standalone tool and as a component of other software we distribute and call GPRO [11], the professional tool tailored for the management of large volumes of data in omic analysis. Currently, TIME 1.0 is available as a first fully functional release that will be maintained with future updates.

ACKNOWLEDGMENTS

LICENSE AND DISTRIBUTION

TIME is commercial software owned and distributed by Biotech Vana S.L at [URL1]. This software is subject to a License Agreement you should accept during installation and may not be copied, reproduced or otherwise transmitted or recorded, for any purpose, without prior written permission from the owner.

INSTALLATION

The application is distributed as an installer for Windows XP/Vista/7 (32 bit and 64 bit), a self-extracting disk image for Mac OS X 10.5 or later (64 bit), and a compressed tarball archive for Linux 2.6 kernel series or later (32 bit and 64 bit).

REQUIREMENTS

TIME requires Java 6 or later. The minimum system requirements for TIME are a PC with a Pentium 4 1.5 GHz or AMD Athlon XP 1500+ processor or higher with at least 1 GB of RAM.

CONCLUDING REMARKS

TIME is a powerful biological sequence editing software displayed in a clean and streamlined interface focused on making most operations one click away. It differs from other similar bioinformatic tools such as Gene Runner [URL 2] in

LITERATURE

1. Pettersson E, Lundeberg J, Ahmadian A: Generations of sequencing technologies. *Genomics* 2009, **93**: 105-111.

2. Bentley DR: Whole-genome re-sequencing. *Curr Opin Genet Dev* 2006, **16**: 545-552.

3. Lundin S, Stranneheim H, Pettersson E, Klevebring D, Lundeberg J: Increased throughput by parallelization of library preparation for massive sequencing. *PLoS One* 2010, **5**: e10029.

4. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA et al.: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005, **437**: 376-380.

5. Schena M, Heller RA, Theriault TP, Konrad K, Lachenmeier E, Davis RW: Microarrays: biotechnology's discovery platform for functional genomics. *Trends Biotechnol* 1998, **16**: 301-306.

6. Wang L, Li P, Brutnell TP: Exploring plant transcriptomes using ultra high-throughput sequencing. *Brief Funct Genomics* 2010, **9**: 118-128.

7. Nowrousian M: Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryot Cell* 2010, **9**: 1300-1310.

8. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: Gen-Bank. *Nucleic Acids Res* 2010, **38**: D46-D51.

9. Kulikova T, Akhtar R, Aldebert P, Althorpe N, Andersson M, Baldwin A et al.: EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Res* 2007, **35**: D16-D20.

10. Futami R, Muñoz-Pomer L, Dominguez-Escriba L, Covelli L, Bernet GP, Sempere JM et al.: GPRO The professional tool for annotation, management and functional analysis of omic databases. *Biotechvana Bioinformatics*: 2011-SOFT3 2011.

URLS

1. **TIME Web Site:** http://www.biotechvana.com/software/time
2. **GENE RUNNER:** http://www.generunner.net/