

 Content archived on 2024-06-18



Algorithms and Tools for Mining Biological Sequence Data

Results in Brief

Algorithms to analyse nucleic acid sequences

Bioinformatics exploits software tools to analyse complex biological data. Improved methods for processing of sequence data related to nucleic acid sequences and proteins overcome current limitations and are supporting sequencing efforts.



DIGITAL ECONOMY



© Thinkstock

Computers have changed the way that fields from industrial automation to quantum mechanics to biomedicine address complex problems. Bioinformatics seeks to establish a pathway linking genetic information (sequences of nucleic acids or proteins) with phenotype (observable characteristics, symptoms or dysfunctions).

The EU-funded research project 'Algorithms and tools for mining biological sequence data' (ALMOND) was launched to develop new techniques for several important problems in computational molecular biology. Methodology emphasised dynamic programming methods that identify simpler 'sub-problems' of a difficult problem, recurrent patterns that relate them and subsequent solution of the base cases.

ALMOND researchers developed novel efficient algorithms for comparison of protein sequences. They focused on a new variant of special alignment path-constrained sequence alignments (called sequence alignment with regular expression path

constraint (SA-REPC)). Researchers delivered two new solutions to this sequence analysis problem and both are available for download from the [host group's website](#) .

Investigators also developed new algorithms for comparison of RNA sequences and structures for the case when the RNA sequences are in a coding region, as commonly occurs in viruses and bacteria. The methods enable prediction of the most likely common ancestor of two RNAs, overcoming the limitations of usual comparison algorithms.

A number of algorithms addressed issues related to next-generation sequencing (NGS). Mapping short reads against an existing reference genome is the first step of many NGS data analyses. New mapping methods outperform existing algorithms, providing substantial improvements. A novel data structure for a graph used by most practical genome assembly methods for NGS data overcomes a major barrier in computational processing of the data. Its 30-40 % increase in memory space is now being exploited in third-party software (Minia).

ALMOND delivered important new algorithms overcoming limitations to current tools addressing bioinformatics and sequence analysis. The above results and others have been published widely and the project has spawned new collaborations between France and Israel. The project is thus expected to have long-lasting impact on the socioeconomically important field of bioinformatics.

Keywords

[Nucleic acid sequences](#)

[bioinformatics](#)

[biological data](#)

[molecular biology](#)

[next-generation sequencing](#)

Discover other articles in the same domain of application



Middleware system utilises AI to coordinate building energy management

8 May 2020



Getting to the bottom of endometriosis with Lucy

26 October 2023



Bio-inspired device slashes power consumption

26 September 2022



Enabling interoperability and seamless data exchange in the energy sector

20 March 2025



Project Information

ALMOND

Funded under

Grant agreement ID: 298761

Project closed

Start date

15 July 2012

End date

14 July 2014

Specific programme "People" implementing the Seventh Framework Programme of the European Community for research, technological development and demonstration activities (2007 to 2013)

Total cost

€ 250 106,00

EU contribution

€ 250 106,00

Coordinated by
BEN-GURION UNIVERSITY OF
THE NEGEV
 Israel

Last update: 5 August 2015

Permalink: <https://cordis.europa.eu/article/id/169407-algorithms-to-analyse-nucleic-acid-sequences>

European Union, 2025