FAST BREAKING PAPERS - 2008

December 2008

PDF

Desmond G. Higgins talks with *ScienceWatch.com* and answers a few questions about this month's Fast Breaking Paper in the field of Computer Science.

Article Title: Clustal W and clustal X version 2.0
Authors: Larkin, MA;Blackshields, G;Brown, NP;Chenna, R;McGettigan, PA; McWilliam, H;Valentin, F;Wallace, IM;Wilm, A;Lopez, R;Thompson, JD; Gibson, TJ;Higgins, DG
Journal: BIOINFORMATICS
Volume: 23
Issue: 21
Page: 2947-2948
Year: NOV 1 2007
* Univ Coll Dublin, Conway Inst Biomol & Biomed Res, Dublin 4, Ireland.
* Univ Coll Dublin, Conway Inst Biomol & Biomed Res, Dublin 4, Ireland.
(addresses have been truncated)

SW: Why do you think your paper is highly cited?

This paper describes a recent version of a computer program called Clustal (actually Clustal W and Clustal X versions 2.0) which is widely used for aligning sets of related protein or DNA (or RNA) sequences together. Much of modern molecular biology revolves around the determination and analysis of sequences, and one of the most commonly used analyses is to compare a sequence to some relatives. This helps you find out what matters in your sequence or in the family of sequences as a whole and is an essential first step in many widely used sequence analysis protocols.

Clustal has been widely used for this task since it was first written by me in Dublin in the late 1980s. Since then, it has had several changes of direction, but it has always been freely available and we put considerable effort into making it user-friendly and able to align large numbers of sequences on personal computers. As a result, it gets widely used as a standard analysis method. Increasingly, it gets used over the Internet, where it runs on large servers such as the one at the European Bioinformatics Institute (EBI).

"Recent advances in sequencing technology have meant that there will be an increasing need for alignment software, capable of handling larger and larger sets of

The current versions are a result of a collaboration between my lab in Dublin and labs at the European Molecular Biology labs in Heidelberg, Germany, and Hinxton, UK, and in Strasbourg, France.

SW: Does it describe a new discovery, methodology, or synthesis of knowledge?

This was release 2.0 of the program and was the result of a major rewrite and reorganization of the code to make it easier to maintain and to develop new features in the future. Most of the new features are, however, invisible to the users. We have had to do this to help us to port the package to the latest versions of the operating

systems on Macs and PCs.

sequences."

We have also had to do this in preparation for the next phase of development, which will hopefully see the program being released in the future with new capabilities and increased capacity and accuracy. This is also the first time that we have made the Clustal alignment server at the EBI, the principal method of access for the program.

SW: Would you summarize the significance of your paper in layman's terms?

The genetic code of the human genome was fully determined about seven years ago. The entire genomes of a range of other species have also been sequenced or are in the process of being sequenced. This presents a major problem for biologists as they try to compare these genomes to each other or to compare different parts of the same genome to each other, in order to understand them.

Our computer program, Clustal, is widely used to help biologists make these comparisons. Specifically it takes sections of DNA or proteins that are related to each other and tries to line them up so that you can see what they all have in common or how they differ. This is an example of what has become known as "bioinformatics," which is the science of using computers to manage and analyze genome information. The most famous and widely used bioinformatics program is the Basic Local Alignment Search Tool (BLAST) which is used to search databases of sequences.

SW: How did you become involved in this research, and were there any problems along the way?

I became interested in this problem in 1987, when I got tired of making multiple sequence alignments by hand, using word processing software. In 1987, a series of papers were published, describing how to do this automatically, and we adapted this to work quickly on PCs, which utilized very little memory.

SW: Where do you see your research leading in the future?

Recent advances in sequencing technology have meant that there will be an increasing need for alignment software, capable of handling larger and larger sets of sequences. This will place increasing demands on the ability of Clustal to align tens or even hundreds of thousands of sequences. There will also be a need for packages that are able to align sequences from different sources and of varying quality and completeness. More pressingly, there is a great need for software to help analyze and visualize relationships within these large data sets.

Professor Des Higgins
UCD Conway Institute of Biomolecular and Biomedical Research
University College Dublin
Belfield, Dublin, Ireland
Web | See also

Related information:

- This paper was also named a Fast Breaking Paper in Computer Science for August 2008 as well as October 2008.
- View an interview with Des Higgins from in-cites.com.

Keywords: Clustal W and Clustal X versions 2.0, sequence analysis protocols, European Bioinformatics Institute, genetic code, human genome, bioinformatics, multiple sequence alignments, Basic Local Alignment Search Tool.

PDF

back to top

2008 : December 2008 - Fast Breaking Papers : Desmond G. Higgins

Scientific Home | About Scientific | Site Search | Site Map