

The Scientist 13[8]:15, Apr. 12, 1999

Previous

Issue Content

Next



Hot Papers

Hot Papers In Bioinformatics

By None

Edited by: Eugene Russo and Steve Bunk

S.F. Altschul, T.L. Madden, A.A. Schäffer, J.H. Zhang, Z. Zhang, W. Miller, D.J. Lipman, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Research*, 25:3389-3402, 1997. (Cited in more than 500 papers since publication)

Comments by **Stephen Altschul**, senior investigator, National Center for Biotechnology Research, Bethesda, Md.

Tools of the trade have become increasingly important in the field of molecular biology, and few if any tools have proven more important than a sequence database search program called BLAST (Basic Local Alignment



David Lipman, Stephen Altschul, Alejandro Schäffer, and Tom Madden

Search Tool). One of several programs used by researchers to search for similarities between newly uncovered DNA or protein sequences and those previously discovered, BLAST has the distinct advantages of being quick and easy to use. It also has capabilities that allow the program to pick out statistically significant sequence similarities.

Investigators from all over the world typically submit newly found sequences to computers at the National Institutes of Health's National Center for Biotechnology Information (NCBI) and have BLAST hunt for matches among stored sequences; NCBI servers perform more than 70,000 BLAST searches a day. Although developed about 10 years ago,¹ the BLAST program got a tune-up about two years ago, as reported in this paper. With more than 8,700 citations to date, the original 1990 BLAST paper is the most cited of the decade, according to the Institute for Scientific Information's *Web of Science*. "If we hadn't reported anything new in the new paper, the mere fact that it was the current citation for BLAST

would have given it a large number of citations," says lead author Stephen Altschul.

There were, however, two significant new features reported in the updated BLAST program. First, a change in the program's basic algorithm allowed it to run three times faster. It also made possible a "gapped" version of the program that produces alignments of DNA or protein sequences containing insertions or deletions. The original BLAST program reported only ungapped alignments, and then strung them together based on statistics rather than explicitly combining them into a single alignment, a feature, says Altschul, that many users didn't like. Although not guaranteed to always find the optimal alignments, programs such as BLAST and FASTA,² a predecessor to BLAST that's still widely used, do find them most of the time and, by taking advantage of a few algorithmic tricks, can do so quickly. According to Altschul, investigators will occasionally get concerned that these programs have missed something and will, as a result, revert to a more thorough but slower algorithm developed in 1981.³

A second innovation, a version of BLAST called Position-Specific Iterated, or PSI-BLAST, enables the program to search for distantly related sequences by constructing a multiple alignment from the pairwise alignments found for a submitted sequence. BLAST makes a protein profile based on this multiple alignment and searches the stored sequence database with this profile. The procedure is then iterated--new profiles are built and new searches performed--until no additional sequences are found to be related.

"Because you have a multiple alignment, you can construct a more sensitive scoring system and detect relationships that are more distant," Altschul explains. "You can look farther away because you can see which are the important residues." Other programs had achieved this previously, but, says Altschul, referring to PSI-BLAST, "I compare it to the Model T. It was not the first out there by a long shot, but it basically opened up this methodology to anyone who wanted to use it by making it totally automatic and very fast." Altschul and NCBI colleague **Eugene Koonin** recently wrote a tutorial for PSI-BLAST.⁴

As for improvements, there haven't been any major advances in sequence database program capability since the 1997 paper, and Altschul doesn't expect any big jumps in technology in the near future--program speed has been all but maxed out. He and his group are, however, working on improving the quality of the alignments and the breadth of statistical assessment capabilities--particularly crucial, Altschul points out, if scientists are to be able to prevent erroneous chance sequence matches from corrupting results.

- S.F. Altschul et al., "Basic local alignment search tool," *Journal of Molecular Biology*, **215**:403-10, 1990.
- W.R. Pearson, D.J. Lipman, "Improved tools for biological sequence comparison," *Proceedings of the National Academy of Sciences*, **85**:2444-8, 1988.
- T.F. Smith, M.S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, **147**:195-7, 1981.
- S.F. Altschul, E.V. Koonin, *Trends in Biochemical Sciences*, **23**:444-7, November 1998.

The Scientist 13[8]:15, Apr. 12, 1999

[Previous](#) [Issue Content](#) [Next](#)

© Copyright 1999, *The Scientist*, Inc. All rights reserved.

We welcome your opinion. If you would like to comment on this article, please write us at [mailto:editorial@the-scientist.com?Subject=The Scientist, Apr. 12, 1999, Hot Papers In Bioinformatics](mailto:editorial@the-scientist.com?Subject=TheScientist, Apr. 12, 1999, Hot Papers In Bioinformatics)

[News](#) | [Opinions & Letters](#) | [Research](#) | [Hot Papers](#) | [LabConsumer](#) | [Profession](#)
[About The Scientist](#) | [Jobs](#) | [Classified](#) | [Web Registration](#) | [Print Subscriptions](#) | [Advertiser Information](#)