Fax NOTE 1612

No. of Pages 3

Today's Date 11 Apr 00

Time

Andy Padilla
DOE

Location

Fax # 510 - 637 - 2018

Telephone #

Comments

From Susan Miller

Company University of Arizona

Location Computer Science

Dept. Charge

Fax #

Telephone #

Original
Disposition:☐ Destroy☐ Return☐ Call for pickup

RECEIVED

APR 14 2000

OSTI

Final Report for ER61911-1004561-0000262**A Flexible Sequence Reconstructor for
Large-Scale DNA Sequencing Projects**Susan Miller
Gene Myers

University of Arizona

Submitted by

Susan G. Miller
11 April 2000

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, make any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

FINAL RESULTS

For the period of January 1999 through August 1999, the FAKtory project was scaled back to a half-time effort for a single systems programmer, Susan Miller. While there had been support for continued development of FAKtory from Sam Pitluck and Bill Kimmerly of LBL and Tom Slezak of LLNL, the JGI declined the use of FAKtory in their sequence assembly process. Sam Pitluck continues to use the FAKII kernel for assembly projects, and is still interested in using FAKtory.

Because no current assembly system produces perfect results on all data, some users prefer to use FAKtory in tandem with other assembly systems, and compare the results to help identify false joins or chimeric sequences. Our gryphon tool can aid in this process by displaying a graphical view of a FAKtory assembly alongside any other .ace format assembly. One of our goals has been to allow interfacing between different systems, and as a result of these efforts, the FAKII kernel is the preferred assembly engine of many GAP4 users worldwide.

During the same period, FAKtory was adopted by the Biotechnology Computing Facility of Arizona Research Laboratories as their primary sequence assembly tool. The BCF provides research support for the University of Arizona community, and has recruited many enthusiastic clients from multiple disciplines, including users from the Laboratory of Molecular and Systematics Evolution (the University's core Sequencing Facility), as well as the departments of Plant Sciences, Entomology, Plant Pathology, Soil, Water, & Environmental Sciences, Neurobiology, Nutritional Sciences, and Ecology & Evolutionary Biology. Parties in the College of Medicine and the College of Pharmacy have also expressed interest in using FAKtory, and are likely to be doing so in the near future.

Recent development of FAKtory has focused on enhancing output capabilities to increase compatibility with other tools, improving the performance and increasing the robustness of the system, and adding features based on user requests. FAKtory can now output entire assemblies in FASTA format, and can produce PostScript file captures of the Layout, Multi-Alignment and Trace views. For compatibility with the GAP4 system, the FAKII Overlap Graph and Assembly objects can now be saved from FAKtory. As an aid to project organization, input and output are automatically directed to a specified project directory which is selectable via a File Selection Box incorporated into the File Directories panel. A File Selection Box is also provided when an output operation is initiated to allow easy redirection of output.

To accommodate the request that very large reference fragments (i.e. over 10Kb in length) be permitted in FAKtory assemblies, fragments that are longer than the pre-compiled limit for the FAKII kernel are detected and entered into the FAKII overlap graph without being subject to the overlap computation. This entry is done before fragments within the length limit are added, so that the latter fragments are overlapped with all fragments. This allows a reference fragment of unlimited length to be included in a FAKtory project without re-compiling or enlarging the FAKII kernel score table.

A visual indicator of multialignment columns that need to be inspected for editing has been added. The speed of jumping from one column to another with traces displayed has been increased. Constraint matchers have been made more general by the implementation of "all-against-all" regular expression matchers, in addition to the existing pairwise matchers. Regular expression matchers may be used to constrain groups of fragments with respect to one another.

Ongoing development of FAKtory is being supported by the Arizona Research Laboratories. The project will be entered into a GNU Public License, and will be available free of charge to interested parties.

Papers and Other Products:

FAKtory: DNA sequence fragment assembly system

Gryphon: DNA sequence assembly comparison tool

E. Myers "Whole-Genome DNA Sequencing," IEEE Computational Science and Engineering, 1, 3 (1999), 33-43.

E. Anson and E. Myers, "Algorithms for Whole Genome Shotgun Sequencing," 3rd Conf. on Computational Molecular Biology (Lyon, France, 1999), 1-9.

E. Myers "A Dataset Generator for Whole Genome Shotgun Sequencing," Conf. on Intelligent Systems for Molecular Biology (Heidelberg, Germany, 1999), 202-210.

Internet sites:

<http://www.cs.arizona.edu/factory>

<http://bcf.arl.arizona.edu>

Other notes:

FAKtory remains in use at several locations worldwide, including University of Regensburg (Germany), and Odense University (Denmark). Many GAP4 users have requested the FAKII assembly kernel for use with the Staden package, and we have cooperated with the software group at the Sanger Centre to provide libraries for various UNIX systems for this purpose.