Educational Module on a HPC Bioinformatics Algorithm

Angela B. Shiflet1, George W. Shiflet1, Daniel S. Couch1, Mario Cannataro2

*1Wofford College, Spartanburg, SC, USA*

*2University “Magna Græcia” of Catanzaro, Catanzaro, Italy*

*shifletab@wofford.edu, shifletgw@wofford.edu, couchds@email.wofford.edu*, *cannataro@unicz.it*

**Abstract**

Prof. Angela Shiflet in computer science and mathematics and Prof. George Shiflet in biology are Fulbright Specialists. In January, 2015, they participated in a three-week collaborative project at University “Magna Græcia” of Catanzaro in Italy, in the Department of Informatics and Biomedical Engineering, hosted by Prof. Mario Cannataro. While there, the three along with Prof. Pietro Guzzi started a project to develop educational module(s) on one or more high-performance-computing bioinformatics algorithms. Drs. Cannataro and Guzzi have written a book, *Data Management of Protein Interaction Networks* (Wiley, 2011), and regularly teach bioinformatics and HPC. Upon returning to the United States, the Drs. Shiflet applied to have undergraduate Daniel Couch be a Blue Waters Intern for one year working on the project. The NSF-funded Blue Waters Project, which provides a stipend for the intern, supports “experiences involving the application of high-performance computing to problems in the sciences, engineering, or mathematics” (http://computationalscience.org/bwsip/). Besides having had an HPC course, the student participated in a two-week workshop at the National Center for Supercomputing Applications (NCSA) facilities on the University of Illinois Urbana-Champaign campus. In the project, he has written sequential and HPC programs and performed timings to accompany an educational module on “Aligning Sequences⎯Sequentially and Concurrently,” available at http://www.wofford.edu/ecs/, and is working with the professors on developing other modules and programs. After covering the necessary biological background, the named module develops the sequential Needleman-Wunsch Algorithm (NWA) to determine the similarity and the alignment(s) that yield a highest similarity score. Employing timings developed by the intern, the module illustrates that the algorithm’s runtime is proportional to the square of the number of nucleotides. Having motivated the need for HPC, the module discusses HPC pipeline versions of NWA along with timings. To aid students, the module contains fifteen Quick Review Questions, many with multiple parts; nine exercises; and five projects. Completed sequential and parallel C with MPI programs are available upon request by instructors. The materials are current being used by students and faculty members a bioinformatics course at University “Magna Græcia” of Catanzaro.

Keywords: Bioinformatics, HPC, undergraduate, computational science, algorithms, Blue Waters, Fulbright