

Lecture on Progress toward Petascale Applications in Bioinformatics and Computational Biology

The 7th IEEE International Conference on Bioinformatics and Bioengineering
Cutting-Edge Research Tutorial Lecture, October 14-17, 2007,
Harvard Medical School Conference Center, Boston, Massachusetts, USA



Craig A. Stewart

Indiana University

Malinda Lingwall

Indiana University

David Bader

Georgia Institute of
Technology

Abstract—

Over the past several years there have been repeated analyses of the potential value of petascale bioinformatics and computational biology applications, as well as analyses of the system engineering steps required to implement applications and systems at such scale. Most recently and notably, Snaveley et al. published the “Workshop Report: Petascale applications in biological sciences”. By one measure the era of petascale computing in biology began in 2006 with the successful clocking of the Riken Institute Protein Explorer system at 1.0 PetaFLOPS. Still, the state of the art of current applications in bioinformatics and computational biology is generally yet orders of magnitude away from petascale, especially in terms of actual performance. The purpose of this lecture is to survey the current state of the art in computational biology and bioinformatics at scale. Suggested topics for papers and posters include, but are not limited to, the following specific subjects: What is the current upper limit of scale of applications in bioinformatics and computational biology? What are the factors limiting scalability of these applications? Can we, as recommended by Snaveley et al., identify candidate petascale applications in any of the following areas: biomolecular structure modeling, modeling complex biological systems, genomics, customized patient care, ecological components of

earth systems modeling, infections disease modeling, or other areas? What are the best ways to measure performance scalability of bioinformatics and computational biology applications? Can we measure what really counts in terms of next generation bioinformatics applications with FLOPS and bytes? The NSF workshop organized by Snaveley, Jacobs, and Bader identified several specific applications as candidates for scaling. The resulting report called for attention to progress in scaling applications by identifying problems, resolving those problems, and trying to anticipate problems at a larger scales and making the step to larger scales. Presentations that discuss the steps, challenges, and solutions to incremental scaling of bioinformatics and computational biology applications are particularly encouraged. Practice and experience papers related to this topic will be of particular value to the scientific community as we strive toward petascale applications.

Bio-

Craig A. Stewart is the Associate Vice President of Indiana University for Research Computing and the University Information Technology Services. He is also Chief Operating Officer of Indiana University Pervasive Technology Labs; Co-Director of Indiana University, Center for Computational Cytomics; and Adjunct Associate Professor of the Department of Medical and Molecular Genetics at Indiana University School of Medicine, Indiana University Purdue University, Indianapolis and Department of Biology at Indiana University, Bloomington main campus. Prof. Stewart received his Ph.D. in Biology from Indiana University in 1988, and has held a variety of leadership positions in Information Technology at Indiana University. He has taught bioinformatics, and his own research activities focus on supercomputer applications in the life sciences.