

Topic 17

High Performance Bioinformatics

Mohammed J. Zaki, David A. Bader, Johan Montagnat, and Concettina Guerra

Topic Chairs

Bioinformatics is the science of managing, mining, and interpreting information from biological sequences and structures. Genome sequencing projects have contributed to an exponential growth in complete and partial sequence databases. Similarly, the rapidly expanding structural genomics initiative aims to catalog the structure-function information for proteins. Advances in technology such as microarrays have launched the subfield of genomics and proteomics to study the genes, proteins, and the regulatory gene expression circuitry inside the cell. What characterizes the state of the field is the flood of data that exists today or that is anticipated in the future. Combined with the fact that many of the bioinformatics tasks are highly compute intensive (e.g., ad initio protein folding), it is clear that high performance computing has a fundamental role to play in various bioinformatics problems ranging from the protein folding problem to large-scale genomics to inferring pathways and regulatory networks. The goal of this session is to present the latest research in high-performance computing applied to bioinformatics tasks. We are especially interested in scalable, parallel, and distributed algorithms for mining and analyzing bioinformatics data, as well as system tools that support large-scale high performance bioinformatics.

There were six papers submitted to this session, out of which three were selected for presentation at the conference. Weiguo Liu and Bertil Schmidt, in their paper, *A Generic Parallel Pattern-based System for Bioinformatics*, present a generic programming framework for sequence alignment algorithms in bioinformatics. Alexandros Stamatakis, Thomas Ludwig, Harald Meier, in *Parallel Inference of a 10.000-taxon Phylogeny with Maximum Likelihood*, present their work on using parallelism to construct the phylogenetic or evolutionary trees for a large number of sequences. Gilles Parmentier, Denis Trystram, Jaroslaw Zola, in *Cache-based parallelization of multiple sequence alignment problem*, use their CaLi software cache library to study a parallel multiple sequence alignment algorithm. These papers represent some of the new advances in the field of high performance bioinformatics, and we hope they will stimulate further research in this exciting area of research.