High Performance Computing for Plant Genomics and Systems Biology

Srinivas Aluru
School of Computational Science and Engineering
Georgia Institute of Technology
aluru@cc.gatech.edu

The invention and rapid proliferation of high-throughput technologies, and accumulation of data in public repositories, create unprecedented opportunities and big data challenges for plant sciences research. High performance computing (HPC) at various scales – high-end supercomputers, moderate size clusters, multicore processors and manycore co-processors – can be effectively utilized to address these challenges. To do so, investments in development of parallel methods, and effective means to deliver them to plant sciences researchers without requiring them to acquire HPC knowledge, are required. In this talk, I will briefly describe current work in our group on developing parallel solutions for next generation sequencing bioinformatics and reconstructing genome-scale networks. I will present our work on reverse engineering genome-scale *Arabidopsis* gene networks from large-scale gene expression measurements, and analyzing them to extract biologically meaningful hypotheses. Despite a decade of progress, constructing such networks for complex multicellular organisms continues to remain a challenge. While simpler models easily scale to large number of genes and gene expression datasets, more accurate models are compute intensive limiting their scale of applicability. I will present our research on the development of parallel mutual information and Bayesin network based methods to eliminate such bottlenecks, and facilitate large-scale network structure learning using high performance computing.