

Demands and Solutions for Genome-Scale Combinatorial Analysis

Mike Langston¹, Elissa Chesler², John Eblen¹, Loren Hauser², Bob Hettich²,
Phil LoCasio², Andy Perkins¹, Arnold Saxton¹, Dave Tabb³, Dorothea Thompson⁴,
Nathan VerBerkmoes², Brynn Voy², Roumyana Yordanova¹, Bing Zhang³, Yun Zhang¹

¹ University of Tennessee, ² Oak Ridge National Lab, ³ Vanderbilt University, ⁴ Purdue University

Abstract. An explosion in computational requirements has occurred as the dimensionality and sheer volume of biological data in need of analysis continues to increase. Resource demands are compounded by data from orthologs, isolates, variants and communities, whose effect is to increase problem size dramatically. Another important factor is the scale of correlation across diverse data types, which tends to lower relevant thresholds and boost edge densities within correlation graph structures. Faster sampling rates, greater numbers of environmental states and other factors combine to enlarge problem size further. Yet the combinatorial problems underlying many central biological questions remain formidable, thereby fueling the immense growth in the demand for petascale computing. Memory bottlenecks are often also highly significant. To address this situation, we have devised a suite of novel mathematical methods and implementations that scale with available resources and reduce exponential growth to the parameter. Putative co-regulation analysis on shared-memory architectures is used as a representative case study.