A comprehensive analysis of *C. elegans* protein kinases from genomic and EST sequence databases

Sucha Sudarsanam and Greg D. Plowman

sucha-sudarsanam@sugen.com

Sugen, Inc.
230 East Grand Avenue
South San Francisco CA 94080

The past few years of gene discovery have been dominated by data mining of cDNA sequences derived through random sequencing of short expressed sequence tags (ESTs). A major limitation of this approach is the biased and incomplete nature of the clones selected for sequencing. Extrapolations are often made as to the total number of genes in a given organism based on analysis of clustered EST databases. However, the accuracy of such calculations remains in question until both comprehensive collection of ESTs and complete genomic sequence is available for a eukaryotic organism.

Availability of the complete genomic sequence of *C. elegans* combined with access to a large number of ESTs, provides a unique opportunity to explore the information content of ESTs. We have recently completed a comprehensive analysis of protein kinases present in the *C. elegans* genome (manuscript submitted). Subsequently, we have compared the diversity of kinases predicted from ESTs with that identified through genomic analysis. We will present our approach, the tools used, and our findings, which highlight the value of both EST and genomic strategies for gene discovery.