

# 5<sup>th</sup> Annual Joint Bioinformatics Workshop

July 19, 2005  
2229 Seamans Center  
University of Iowa

by

**Annie Chiang**  
University of Iowa

## ***Computational Orthologous Prioritization (COP): A comparative genomic approach toward candidate gene prioritization for disease gene identification***

### ***ABSTRACT***

In disease gene discovery efforts, the selection of candidate genes for mutation screening is an evolving process that relies on available, accumulating data that is typically disease-specific. Prioritization efforts may be imprecise and inconsistent when based upon conclusions drawn from ever accumulating evidence and the scientific literature, since different conclusions may be derived from different subsets of data. It is believed that disease-causing mutations are more likely to occur at conserved sites, as measured by comparisons of orthologous sequences across a diverse array of organisms. This is based upon the notion that conserved amino acids are more likely to be functionally important. Mutations at these functionally important sites, therefore, have a higher probability of resulting in the disease phenotype. With the availability of various genome sequences produced concurrently with that of the human genome, comparisons at a global scale may provide additional insight toward better understanding of the genetics of human disorders. Building upon the belief that disease-causing mutations tend to occur with higher probability at conserved sites, we describe a comparative genomic approach, Computational Orthologous Prioritization (COP), that prioritizes candidate genes based on sequence conservation. The COP approach prioritizes candidate genes based on the sequence similarities (positive subset) or dissimilarities (negative subset) across organisms. Genes with high sequence similarity to genes found in the positive subset are ranked higher while genes with high sequence similarity to genes in the negative subset are ranked lower. The fundamental basis of the COP method is that species sharing conserved structural and/or functional traits, tend to share genes that underlie the biological processes responsible for these features. Similarly, species that do not share these features are less likely to contain these genes. Additionally, those genes that are involved in the parallel structural and/or functional processes are more likely to be evolutionarily conserved and therefore orthologous to one another. Thus, using the presence or absence of biological feature(s) as a criterion, one can readily identify two subsets of mutually exclusive species that can serve as positive and negative filters. This method does not preclude the use of other prioritization methods (such as functional annotations) but rather is meant to augment the existing disease gene identification process. The COP method has been successfully applied toward the correct identification of a Bardet-Biedl Syndrome (BBS3) gene. We hypothesize that the COP approach may be employed in the study of other genetic disorders.