**COLLOQUIUM**

Department of Computer Science and Engineering

University of South Carolina

### **On the Edit Distance between Genomes**

###  **with Duplicate Genes**

### **Mingfu Shao**

### Laboratory of Computational Biology and Bioinformatics

### École Polytechnique Fédérale de Lausanne

Date: **March 28, 2014**

Time: **1430-1520 (2:30pm-3:20pm)**

Place: **Swearingen 2A15**

# Abstract

A basic problem in comparative genomics is to compute the edit distance between two genomes, i.e., the minimum number of large-scale evolutionary events that can transform one genome into the other. These evolutionary events fall into two categories: rearrangements and content-modifying operations. Most of the genomic rearrangements events can be modelled by the double-cut-and-join (DCJ) operation, which has formed the basis for much algorithmic research on rearrangements over the last few years. Content-modifying operations include insertions, deletions, and segmental duplications. In this talk, we first discuss the edit distance in terms of DCJ operations and segmental duplications. A new graphical data structure, the trajectory graph, is introduced to model and possibly refine any given evolutionary trajectory between two genomes. We then show how to compute the edit distance between genomes in the presence of duplicate genes. We formulate it as an integer linear programming problem and provide an efficient preprocessing approach while preserving optimality. To the best of our knowledge, this is the first practical approach to compute the exact edit distance between genomes with duplicated genes, which also results in more accurate assignments of orthologs between genomes than state-of-the-art methods.

**Mingfu Shao** is a third-year Ph.D student in the Laboratory of Computational Biology and Bioinformatics, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland. His research interests focus on comparative genomes, especially designing algorithms for problems arising in genomic rearrangements and phylogeny construction. Mingfu received his BS degree from the Institute of Beijing Technology, and his MS degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China