



ALGORITHM FOR RECONSTRUCTING THE LIFE HISTORY OF TUMORS FROM GENOMICS DATA

Research Challenge

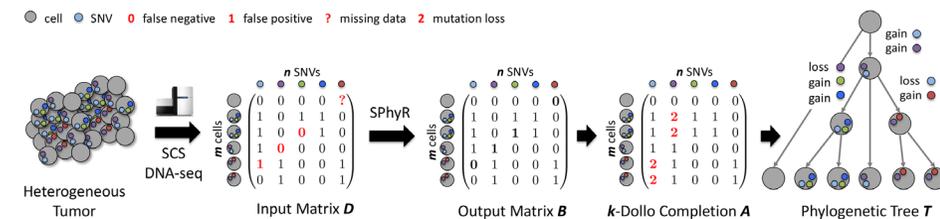
The phylogeny estimation problem from single-cell sequencing data is a variant of the classic phylogeny estimation problem with incorrect and missing data due to the used sequencing technology. Current methods aim to simultaneously construct a phylogenetic tree and correct these measurement errors using either too stringent or too permissive evolutionary models. There is a need for methods that employ appropriate evolutionary models that strike a balance between being realistic and yet, sufficiently constrained.

Methods & Codes

To solve the phylogeny estimation problem from single-cell sequencing data, the researchers used techniques from combinatorial optimization. More specifically, they formulated the problem as an integer linear program (ILP) and designed and implemented a custom column generation and cutting plane approach for the ILP formulation.

Why Blue Waters

The computational resources of Blue Waters allowed the research team to perform these experiments at scale, enabling them to study the performance of their algorithms and the underlying problem statements in many different experimental settings.



Tumors are composed of cellular populations with distinct sets of mutations. Single-cell sequencing of a tumor yields an input matrix D with incorrect and missing entries. SPhyR aims to simultaneously correct errors in matrix D and infer the evolutionary history of the cells.

Results & Impact

In light of frequent loss of point mutations in cancer due to copy-number aberrations, the k-Dollo model is more appropriate than the evolutionary models utilized by previous methods. This project resulted in a novel combinatorial characterization of solutions to the underlying computational problem as constrained integer matrix completions, which formed the basis for the efficient integer linear programming approach utilized by SPhyR.