Evolutionary costs in gene-species reconciliation

Introduction

Scoring differences between gene and species trees is indispensable for evolutionary studies. While many gene trees are unrooted and species trees are typically rooted, standard scoring tools do not accept such inputs. We present a novel framework that allows to compare rootings of the unrooted gene tree with the species trees by using rooted scoring functions such as Robinson-Foulds (RF), gene duplications (D), gene losses (L), duplication-losses (DL) and deep coalescence (DC).

Cost examples

RF cost example. Each edge of the unrooted gene tree $G$ is labeled by the RF score between the species tree $S$ and the rooting of $G$ placed on this edge. There are 7 optimal rootings forming a “valley” with the RF score 12.

DL cost example with the star-like topology for the trees from the previous figure. There are 3 optimal edges having cost 20.

Gene trees and star topologies

Transforming a gene tree into a star topology. Required: lca-mappings from the rooted subtrees of the input gene tree into a given species tree $S$.

Properties of stars and optimal regions

- Star topology of a given gene tree depends on lca-mappings only.
- Each double and empty edge is optimal for all five costs.
- All edges of star M6 are optimal.

We have the following relationships between optimal regions of gene trees under our cost functions: $D \supseteq DL = L \subseteq DC \subseteq RF \{1,2,3\}$.

Given unrooted gene tree $G$ and a rooted species tree $S$, the optimal “unrooted” cost (RF, DC, DL, L and D) as well as the costs of all rootings of $G$ can be computed in $O(n)$ time, where $n$ is the size of input trees $\{1,2,3\}$.

Applications: improving supertree heuristics

Gene tree parsimony (GTP) problems: given a collection of gene trees and a cost function find an optimal species tree that minimizes the total cost.

GTP problems are usually NP-complete. They often are addressed by local search heuristics that perform a stepwise search of the tree space, where each step is guided by an exact solution to an instance of a local search problem.

Local search problem (NNI): given an unrooted gene tree $G$ and a rooted species tree $S$ find an optimal species tree $S'$ in the 1-NNI neighborhood of $S$.

Fast local search algorithm

Reconciliation algorithm based on star topologies can be adopted to design fast local search algorithms.

References