

SEADOG v1.0 Manual

Description

SEADOG (short for "Simultaneous Evolutionary Analysis of Domains and Genes through phylogenetic reconciliation") is a software package for simultaneous inference of domain-level and gene-level evolution through a joint phylogenetic reconciliation of domain, gene, and species trees. The software takes as input a rooted or unrooted domain tree, rooted gene trees for the gene families in which the domains of the domain tree occur, and a rooted species tree on the species considered in the analysis. It outputs a reconciliation of the domain tree with the gene trees and of the gene trees with the species tree. SEADOG implements the Domain-Gene-Species (DGS) reconciliation model and heuristic algorithm introduced in the following manuscript:

[An Integrated Reconciliation Framework for Domain, Gene, and Species Level Evolution](#)

Lei Li and Mukul S. Bansal

Asia-Pacific Bioinformatics Conference (APBC 2018), and

IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB); in press.

SEADOG is available open source under GPL version 3. The software is free to use, but WITHOUT any guarantee of correctness.

Input

The required input consists of one domain tree, a set of gene trees, and a species tree. All trees should be in Newick format, without any labels for internal nodes, and should be rooted and binary. It is possible to have SEADOG treat the domain tree as unrooted using the -u command line option (see below), but the tree itself should be written as rooted and binary in the domain input file. Each tree must appear in a separate file. The domain tree and species tree can have arbitrary file names, but gene tree names should be of the form "geneTreeName.tree". All gene trees should be in a single directory.

Each leaf label (domain name) in the domain tree must be of the form "domainName_GeneID_GeneTreeName", where the domainName is any label for that specific domain sequence, GeneID is the unique gene ID or name of the specific gene that contains that domain sequence, and GeneTreeName is the name of the gene tree that contains that specific gene. For example, "domainXYZ_FBgn0100324_geneTree4".

Likewise, gene names in the gene trees should be of the form "GeneID_SpeciesName", where GeneID is the unique gene ID or name for that gene, and Species name is the label of the species from which that gene was sampled.

Command Line Arguments

The program takes the following command line arguments, among which -d, -g, and -s are required and the others are optional.

- u: Assume the domain tree is unrooted and consider all possible rootings of the domain tree to find one that minimizes the DGS reconciliation cost.
- d The input domain tree file.
- g Path to the directory containing the gene trees.
- s The input species tree file.
- o Output file name. By default the output file will be the input domain tree file name plus ".output".
- DD Domain duplication cost. Default value 2.
- DL Domain loss cost. Default value 1.
- DTA Domain transfer cost when the donor and recipient are in the same gene family. Default value 4.
- DTB Domain transfer cost when the donor and recipient are in different gene families. Default value 6.
- GD Gene duplication cost. Default value 2.
- GL Gene loss cost Default value 1.

Command Line Example:

```
./SEADOG.linux -d domaintree.tree -g GeneTrees/ -s 12flies.stree -o output.txt
```

Output

The reconciliation output begins with a listing of the domain tree, gene trees, and species tree, with their internal nodes labeled. Each internal node is labeled by an index, followed by the indices of its two children.

The next output block shows the reconciliation of the domain tree with the gene trees and shows the event type and mapping for each domain tree node. The reconciliation between the gene tree(s) and species tree appear next, showing the event type and mapping for each node in the gene tree(s).

The output reconciliation ends with three lines showing the final DGS reconciliation cost, the reconciliation cost between the domain tree and gene trees, and the reconciliation cost between the gene trees and species tree, respectively.

Contact

If you have any questions, suggestions, or concerns, please contact either Lei Li (lei.li@uconn.edu) or Mukul Bansal (mukul.bansal@uconn.edu).