Software for isometric gene tree reconciliation

DOMINIKA MIHÁLOVÁ

SUPERVISOR: DOC. MGR. BRONISLAVA BREJOVÁ, PHD.

Progress

First chapter

- ◦ Introduction to the topic

- ◦ Explanation of basic terms, essential terminology from bioinformatics; gene tree, species tree as trees from graph theory

- ◦ Definition of gene tree reconciliation – mapping gene tree to species tree

- ◦ Introduction of existing software with different gene tree reconciliation algorithms; three types:
  - ◦ Scoring gene tree reconciliation
  - ◦ Probabilistic gene tree reconciliation
  - ◦ Isometric gene tree reconciliation

- ◦ GitHub of project: https://github.com/mihalova/diploma_thesis

Software

- Goal: software for isometric reconciliation of unrooted gene tree and rooted species tree

- Source code based on master's thesis of Radoslav Chládek (Algorithms for isometric gene tree reconciliation. Univerzita Komenského v Bratislave, 2019)

- Programming language: Java (SDK 11)

- Libraries: lpsolve (linear programming solver – used in reconciliation algorithm), junit-jupiter-api (fro running unit tests)

My work

◦ Changing parser to parse unrooted gene tree with intervals or add intervals to gene tree

◦ Parse mapping of leaves from gene tree to species tree

◦ Separating one Reconciliator class to three (Main, Printer and Reconciliator)

```java
if (tolerance != null && !time.contains("-")){
    //nastavenie tolerance pre hrany
    double t = Double.parseDouble(time);
    double[] interval = makeInterval(t, tolerance);
    min_time = interval[0];
    max_time = interval[1];
} else if (time.contains("-")) {
    //parsovanie intervalov 2.02-3.01 = double "-" double
    String[] interval = time.split( regex: "-");
    min_time = Double.parseDouble(interval[0]);
    max_time = Double.parseDouble(interval[1]);
} else
    throw new IllegalArgumentException("Insert tree with intervals or set tolerance.");


str = str.substring(0, colonIndex);
if(!str.startsWith("(")){
    //tu bude parsovanie leafmap G0_S1 = nodeG string "_" nodeS string
    String[] mappingG  = str.split( regex: "_");
    G.addMapping(str, mappingG[1]);
```

Input and output

Input: (path to directory with trees and tolerance - optional)

1. Path to directory with rooted species tree and unrooted gene tree with intervals
2. Path to directory with rooted species tree, unrooted gene tree with intervals on some edges and tolerance
3. Path to directory with rooted species tree, unrooted gene tree without intervals and tolerance

Output: most parsimonious reconciliation of gene tree to species tree (Radoslav Chládek's algorithm)

Example of reconciliation

```
Total DL of solution: (3,1)
Number of the most parsimonious solutions: 1
Minimal DL: (3,1)

#[min depth] [max depth]
root 0.0162725  0.07937
G11_S20G12_S21 0.13371  0.17307
G11_S20 0.2301  0.2301
G12_S21 0.23484  0.23484
G0_S15G1_S16G3_S18G4_S18G5_S18G7_S18 0.110875  0.13432
G0_S15G1_S16 0.19174  0.19917
G0_S15 0.22723  0.22723
G1_S16 0.22835  0.22835
G3_S18G4_S18G5_S18G7_S18 0.1125  0.139195
G3_S18 0.23102  0.23102
G4_S18G5_S18G7_S18 0.13721  0.16969
G4_S18G5_S18 0.22064  0.224865
G4_S18 0.23102  0.23102
G5_S18 0.23102  0.23102
G7_S18 0.23102  0.23102
```

Next work

Software finds the most parsimonious rooting of unrooted gene tree considering to species tree using different isometric reconciliations (different algorithms)

Validation:

◦ Comparison of results from software with considered branch lengths/without considered branch lengths

◦ How much good/bad was the result?

◦ Dependency of the result on changed parameters (length of branch intervals, tolerance of edge)

Bibliography

[1] Bastien Boussau et al. Genome-scale coestimation of species and gene trees. Genome Research, 23(2):323330, Feb. 2013.

[2] Broňa Brejová et al. Isometric gene tree reconciliation revisited. Algorithms for Molecular Biology, 12(1):17, Jun. 2017.

[3] Dave Danicic et al. Notung 2.8: A Manual. Notung Development Team, Mar. 2015.

[4] Jean-Philippe Doyon, Cedric Chauve, and Sylvie Hamel. Algorithms for exploring the space of gene tree/species tree reconciliations. In Nelson C.E., Vialette S. (eds) Comparative Genomics. RECOMB-CG 2008. Lecture Notes in Computer Science , volume 5267, pages 113. Springer, Berlin, Heidelberg, Oct. 2008.

[5] Joseph Felsenstein. Inferring Phylogenies. Sinauer, 2003.

[6] Morris Goodman et al. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Systematic Zoology, 28(2):132163, Jun. 1979.

[7] Damir Hasi¢ and Eric Tannier. Gene tree species tree reconciliation with gene conversion. Journal of Mathematical Biology, 78(6):19812014, May 2019.

[8] Jian Ma et al. The innite sites model of genome evolution. Proceedings of the National Academy of Science, 105(38):1425414261, Sep. 2008.

[9] Matthew D. Rasmussen and Manolis Kellis. A bayesian approach for fast and accurate gene tree reconstruction. Molecular Biology and Evolution, 28(1):273 290, Jan. 2011.

[10] Benjamin Vernot et al. Reconciliation with non-binary species trees. Journal of Computational Biology, 15(8):9811006, Oct. 2008.

[11] Albert J. Vilella et al. Ensemblcompara genetrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Research, 19(2):327335, Feb. 2009.

[12] Wu Yi-Chieh et al. Treex: Statistically informed gene tree error correction using species trees. Systematic Biology, 62(1):110120, Jan. 2013.