**Practicum 4 WORKSHEET Due: 4pm, Tues, Dec. 8th**

**Your name:**

Hand in your assignment by email to [comp-bio@cs.cmu.edu](mailto:comp-bio@cs.cmu.edu) or in MI646 between noon and 4:00pm on Dec.8th.

The goals of this assignment are to give you experience interpreting a gene tree in the context of a species tree and to introduce you to a software package, Notung, that supports this sort of analysis. The tasks are designed to give you skills that will be helpful in future phylogenetic analyses, including your course projects.

Notung juxtaposes a gene tree and a species tree to infer events in the history of the gene family. If a gene family evolves solely through substitution and small insertions and deletions, the gene family tree will agree with the species tree. Larger scale events, including gene duplication, gene loss and horizontal gene transfer, result in a gene tree that disagrees with the species tree. *Reconciliation* is the process of fitting a gene family tree to a species tree to infer (a) the correspondence between each ancestral gene and an ancestral species and (b) the events that occurred in the history of the gene family.

**WHAT TO HAND IN:**

* This worksheet

***Steps 2 - 4***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table I** | | | | |
|  | **Score** | **D** | **T** | **L** |
| **Reconciled tree (DL)** |  |  |  |  |
| **Resolved tree (DL)** |  |  |  |  |
| **Rearranged tree (0.75)** |  |  |  |  |
| **Rearranged tree (0.90)** |  |  |  |  |

**Step 2: *Record the following information in the Practicum 4 worksheet:***

* ***The species associated with the polytom****y:*
* ***The species associated with each of its children****:*

***Step 3: Record the following information in the Practicum 4 worksheet:***

* ***The species associated with the resolved node:***
* ***The species associated with each of its children****:*

***Based on this information, and information that you recorded in step 2, which node is the root of the subtree resulting from the merger?***

***Step 5:***

***Multiple optimal solutions:* *Record the number of optimal solutions in your worksheet.***

***Toggle back and forth between the two histories and record the two labels on the transfer associated with the alternate histories in Table II***

|  |  |  |
| --- | --- | --- |
| **Table II** | | |
|  | **Donor species** | **Recipient species** |
| **Reconciliation (DTL)** |  |  |
| **Reconciliation (DTL)** |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table III** | | | | | |
| **Soln** | **D** | **T** | **L** | **Score** | **Species at the root** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **3** |  |  |  |  |  |
| **4** |  |  |  |  |  |
| **5** |  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table IV** | | | | | |
|  |  | **Transfer at root** | | **Transfer at child of root** | |
| **Soln** | **Transfer at root?** | **Donor** | **Recipient** | **Donor** | **Recipient** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **3** |  |  |  |  |  |
| **4** |  |  |  |  |  |
| **5** |  |  |  |  |  |

***Tables II – IV provide information the optimal solutions recovered by this reconciliation. How many rows are there in Table II? How many rose are there in Tables III and IV? What is the relationship between the number of rows in the tables and the total number of optimal solutions reported at the bottom of the task panel?***

**Step 6: Interpreting the history of the PylB family in light of your analysis**

* Consider the number of events inferred in Steps 2- 4 of your analysis, which you recorded in Table I.
  + ***What impact did rearrangement have on the number of duplications? On the number of losses? Was the change in the number of events similar for both duplications and losses, or was the impact greater on one of them? How do you explain what you observe?***
  + ***What impact did switching to a model with duplications and transfers have on the number of duplications? On the number of losses? Was the change in the number of events similar for both duplications and losses, or was the impact on one of them greater? How do you explain what you observe?***
* Given duplications and losses only, there is only one most parsimonious reconciliation of a gene tree and a species tree. However, when the history of the gene family includes transfers, there may be more than one most parsimonious history of duplications, losses and transfers that explains the disagreement between a gene tree and a species tree. In fact, you recovered several optimal solutions for the trees in this practicum.

Equally parsimonious histories can differ in several ways. Two histories may have the same number of events of each type, but those events may occur in different taxa. Since both histories have the same numbers of events of each type, they must have the same event score. It is also possible to have two histories with the same event cost, but different numbers of events of each type.

**Considering Table II only:**

* **When you click on the green circle closer to the leaves to inspect a different optimal event history,**
  + **does the number of transfers in the subtree rooted at that green circle change?**
  + **does the number of duplications in the subtree rooted at that green circle change?**
  + **does the number of losses in the subtree rooted at that green circle change?**
  + **do the events occur in different locations in the *gene* tree in the two histories?**
  + **do the events occur in different locations in the *species* tree in the two histories?**

**Considering Tables III and IV only:**

* + ***How many optimal histories listed in those tables had the same number of events, but in different locations in the tree?***
  + ***Did you observe any histories where the number of events of each type changed?***
  + ***If so, which events types increased and which events types decreased? Explain what you observe in terms of the reconciliation costs?***
* When there is more than one most parsimonious history, it can be difficult to interpret the results. In this case, it is often helpful to identify common and distinct features of the optimal histories:
  + ***How many histories had a transfer at the root of the tree?***
  + ***How many histories involved more than one transfer between bacteria and archaea?***
  + ***How many histories involved more than one transfer How many histories involved more than one transfer within the Archaea?***
  + ***How many histories imply that the pyrrolysine originated in Archaea? Of those, do they agree on the taxonomic group within the Archaea?***
  + ***How many histories imply that the pyrrolysine originated in Bacteria? Of those, do they agree on the taxonomic group within the Bacteria?***
  + ***Which history do you find most convincing? Why?***
* Pyrrolysine is found in three distantly related groups of bacteria belonging to different Bacterial phyla. It is primarily observed in the Firmicutes and the Detalproteobacteria, but has also been found in one Spirochaete species.
  + ***Does the taxonomic group that is most closely related to the Spirochaete belong to the Firmicutes, the Deltaproteobacteria or the Archaea?***
  + ***The history of events in the clade comprising the Spirochaete and its sister taxon is the same in all of the most parsimonious histories reconciliations. What are those events? Does this event history seem plausible to you or did you expect a different event history for the Spirochaete? Why or why not?***