Comparative and Functional Genomics. Practical session: Phylogenomics

IMPORTANT_1: Submit your answers with your ESCI-UPF gmail account. You can only submit your answers once.

IMPORTANT_2: Wednesday 22th of May, 2024, the last day to submit your answers.

laia.barcenilla@alum.esci.upf.edu

Canvia de compte

L'esborrany s'ha desat

* Indica que la pregunta és obligatòria

Adreça electrònica *

☑ Registra **laia.barcenilla@alum.esci.upf.edu** com el correu electrònic que s'inclourà a la meva resposta

In the discipline of **phylogenomics**, orthology prediction is not the only reason why we work and process large collection of trees. In particular, by reconstructing and comparing the topologies shown by the different gene trees and the species tree, we can infer the occurrence of evolutionary events of interest (e.g., the expansion of a set of gene families in a given ancestral node by means of gene duplications), or we can test evolutionary hypotheses based on the distinct topologies shown by the gene trees.

Go to the exercise folder ('**CFG_PS_phylogenomics24.zip**'). In there you will find a collection of 24 rooted gene trees grouped in a single PDF file ('**geneTrees.pdf**', the tree in pg. 1 corresponds to gene tree 1, the tree in pg. 2 to gene tree 2, etc.). These gene trees have been colored to facilitate their inspection based on the species genome to which each sequence corresponds. The species tree has also been colored accordingly ('**speciesTree.pdf**'). It is usual in these type of analysis to represent species with a certain code either at the beginning or at the end of the sequence identifier (in this case at the end; e.g., sequence Phy003LDJI_FRAVE belongs to the species FRAVE).

Answer the following questions below.

IMPORTANT!: For each question, indicate, in a **numeric sorted list** (from the **lowest to the highest** number), **separated by commas**, which trees satisfy the inquired condition in each case. For example, if gene trees numbers 4, 5 and 8 satisfy a given inquired condition (e.g., gene trees showing species-specific duplications), then the correct answer in this hypothetical case would be "4,5,8".

Answers with **not properly sorted numbers will be incorrect** (e.g., "8,4,5" or "5,4,8").

Avoid also white-spaces in the middle of the numeric list (e.g., "4, 5, 8" would be incorrect in this hypothetical case).

Q1) Based on the species tree, **reproduce the clade** containing PYBRE, PRUAV, PRUMU, PRUDU, MALDO and PRUVE **in Newick format**. (Hint: a reasonable way to test whether your Newick format string is correct is to save it as a text file, open it with a tree visualizer software (e.g., FigTree, or http://phylo.io/, and check whether the expected topology is being displayed).
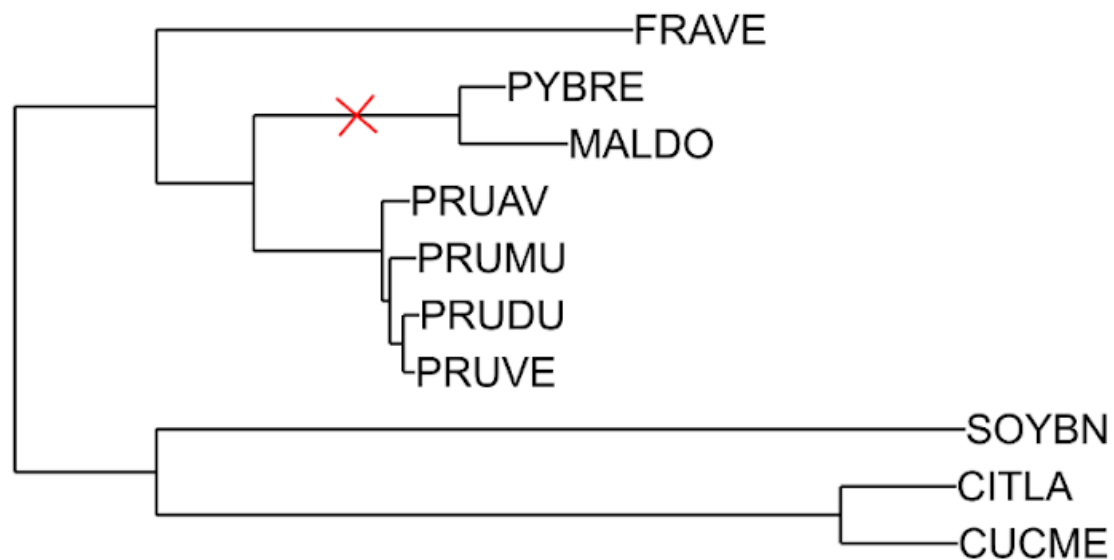
((((PRUDU,PRUVE),PRUMU),PRUAV),(PYBRE,M

Q2) Based on the species-overlap algorithm, **list three gene trees** that experienced **ancestral duplications but not species-specific** duplications?

4,20,21

Q3) Given the species tree and each gene tree reconstruction, which gene families experienced **ancestral duplications** <u>in the node with a red cross</u>?

1,2,3,6,7,9,10,12,13,16,17,19,23,24

Q4) If we assume that species **FRAVE is distantly related from species PRUVE**, which gene tree/s could be showing **horizontal gene transfer** event/s **between these two species**?

2

Q5) Which trees show **some sequences** from PRUMU and PRUAV **forming a monophyletic group** (**monophyletic group = clade**)

3,5,6,8,11,12,18,23

Q6) A gene tree can be rooted at any branch, but not all the roots are equally parsimonious. For this analysis, trees were rooted based on the species tree, but there were some trees for which the optimal root was not so obvious and presented other **equally plausible root options**. **Which trees are these**?

5,8,11,14,23

Q7) There is a gene tree including only sequences from a single species. **Let's assume that this gene tree was reconstructed after searching for similar sequences** in all **currently available genomes in online databases**. How sure we are that this gene family really originated in PRUDU and constitutes an innovation of this species? Indicate **which of the following statements are correct, if any**

○ We can assume with an absolute confidence that this gene family is only found in PRUDU as we have not found homologs in other species

○ We can not discard that this gene family could have been transferred from Bacteria

○ The fact that this gene family includes a lot of copies for this species is a clear evidence that it must be older than expected and is probably present in other species

⦿ Based on currently available data, it seems plausible to consider that these PRUDU sequences do not have orthologs

Esborra la selecció

Q8) See the two alternative topologies below, we are interested in knowing how many gene trees support with no ambiguity each of the topologies shown below:
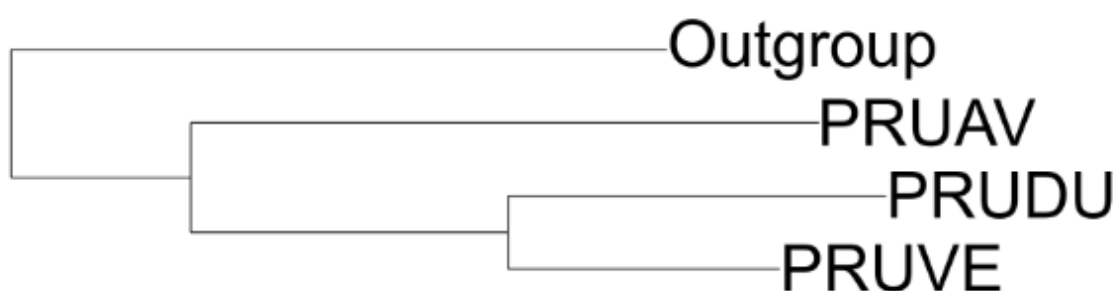
Q8.1) Trees supporting topology A:
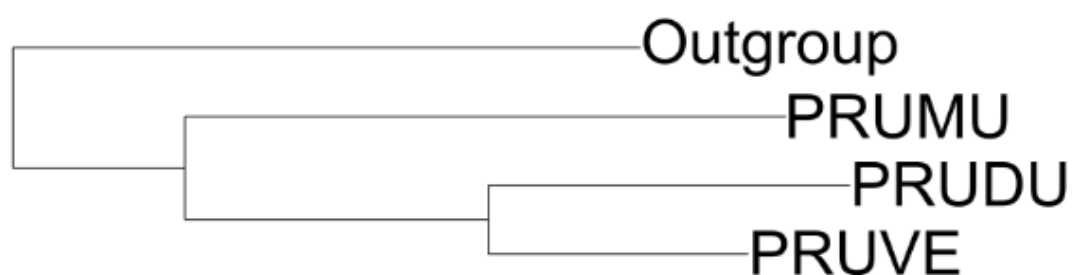
La vostra resposta

Q8.2) Trees supporting topology B:

La vostra resposta

Q9) Find <u>four</u> **identical** trees to the species tree

La vostra resposta

Q10) Which of the following trees are **not identical but congruent** with the species tree?

☐ 3

☐ 7

☐ 14

☐ 15

Q.11) One of the main applications of phylogenomics is to build a species tree based on a set of **gene trees showing suitable properties for being used as phylogenetic markers**. Which trees would have a high consensus in being suitable markers for such analysis?

☐ 9

☐ 23

☐ 18

☐ 15

Finally, we are going to practice again the **species overlap** and the **gene-tree/species-tree reconciliation** algorithms.

Q12) Look at the gene tree shown in '**Phy00CS519_PRUDU_tree.pdf**' (alternatively, you can visualize the gene tree file by uploading 'Phy00CS519_PRUDU_tree.nw' into http://phylo.io/). Count the number of orthologs and paralogs of **Phy00CS519_PRUDU** according to (1) **species overlap** and also according to (2) gene-tree/species-tree **reconciliation**. (For the reconciliation analysis, use the **same species tree** as the one used for previous exercises).

After counting orthologs and paralogs found by each of the two methods, choose the correct answer below:

○ Species overlap and reconciliation found the same number of orthologs

○ Reconciliation found fewer orthologs than species overlap

○ Reconciliation found fewer paralogs than species overlap

○ None of the above is true

Envia   Esborra el formulari

No enviïs mai contrasenyes a través de Formularis de Google.

Aquest formulari s'ha creat fora del vostre domini. Informa d'un ús abusiu - Condicions del Servei - Política de privadesa

Google Formularis