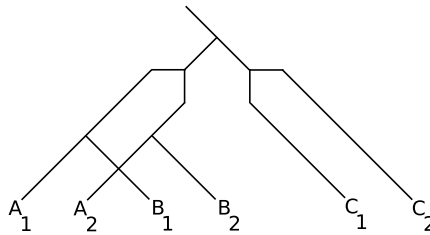# Exercises – Phylogenetics

Exercise Sheet 2 — 19.10.2017

Due: 26.10.2017

Task 1 Gene and species trees. (2 points)

(a) Write down all pairs of paralog and ortholog genes.

(b) Find out what *xenolog* means. Draw a picture and explain the term.

Task 2 Gene tree / species tree reconciliation. (3 points)

Consider the following gene tree G and species tree S.

$$G \;=\; ((g_4, g_5)g_2, (g_6, g_7)g_3)g_1;$$
$$S \;=\; (A, (B, C)n_2)n_1;$$

Let g_4 be present in species A. Genes g_5 and g_6 are in B and gene g_7 is found in C.

(a) Draw G and S. Label the internal nodes and leaves. Furthermore label the leaves of G with the corresponding names of the species.

(b) Calculate $\gamma(g_1)$, $\gamma(g_2)$ und $\gamma(g_3)$. Calculate the $M(.)$ for every node (including the leaves) in G and draw the matching arrows (lecture notes: Figure 2.6) from G to S. Decide for every internal node whether it is a duplication or speciation.

(c) Draw the received scenario from the species tree (similar to the picture in the lecture notes. Figure 2.5, right)
Hint The scenario might include deletions.

Task 3 Last common ancestor. (2 points)

Consider a rooted tree $G = (V, E)$. How can you find the *last common ancestor* (LCA) for two given nodes u and v? Write an algorithm in *pseudo code* that determines $LCA(u,v)$. What is the *worst case* run-time of your algorithm?

Hint: You can find examples for pseudo code in the lecture notes (page 13, 22, 23, 28, 35, 36 etc.)

Don't get confused! *lowest* and *last common ancestor* are basically the same. In phylogenetics, the term *last common ancestor* is more common since we're looking at the evolutionary time of events. However, if you think about vertices and trees in graph theory, the term *lowest common ancestor* is used since there is no time involved in the analysis. Luckily, the abbreviations of the words are the same.