

11. ÜBUNG "BIOINFORMATIK", SS 16

Aufgabe 1: (5 Credits)

Show that the two definitions for cographs are indeed equivalent:

Def 1:

- K_1 is a cograph.
- The disjoint union of two cographs is a cograph.
- The complement of a cograph is a cograph.

Def 2:

- K_1 is a cograph.
- The disjoint union of two cographs is a cograph.
- The join of two cographs is a cograph.

Aufgabe 2: (5+5+5=15 Credits)

Let A, B, C, D be four different species from which we extracted some genetic material. In particular, we observed two independent gene families G^1 and G^2 consisting of the respective subsets G_X^1 and G_X^2 for each species $X \in \{A, B, C, D\}$.

Assume for the first family that $G_A^1 = \{a_1, a_2\}$, $G_B^1 = \{b_1, b_2, b_3\}$, $G_C^1 = \{c_1\}$, $G_D^1 = \{d_1\}$, and the second family that $G_A^2 = \{a'_1, a'_2\}$, $G_B^2 = \emptyset$, $G_C^2 = \{c'_1, c'_2\}$, $G_D^2 = \{d'_1, d'_2\}$.

Using multiple sequence alignments we obtained the following (symmetric) similarity scores for the genes in G^1 and G^2 :

G^1	b_1	b_2	b_3	c_1	d_1	G^2	c'_1	c'_2	d'_1	d'_2
a_1	-1	-1	-2	-2	-3	a'_1	-2	-3	-4	-4
a_2	-2	-2	-1	-2	-3	a'_2	-3	-2	-4	-4
b_1				-2	-2	c'_1			-2	-2
b_2				-2	-2	c'_2			-2	-2
b_3				-2	-2					
c_1					-1					

For simplicity, no two genes from different gene families will be estimated as orthologs or paralogs. Moreover, no horizontal gene transfer occurred.

- Based on the similarity scores determine the estimated orthology relation $\hat{\Theta}_1$ and $\hat{\Theta}_2$ (for both gene families G^1 and G^2 independently) and show that $\hat{\Theta}_i$ ($i = 1, 2$) is a "valid" orthology relation.
- Determine the event-labeled gene tree for each gene family.
- Determine a possible species tree where the two gene trees evolved along and explain your result.

Deadline: Monday - July 11th, 2016 - 2pm