Exercises – Phylogenetics

Universität Bielefeld, WS 2016/2017,
M. Sc. Tizian Schulz

https://gi.cebitec.uni-bielefeld.de/Teaching/2016winter/Phylogenetik

Exercise Sheet 5 — 29.11.2016

Due: 6.12.2016

Task 1 Uninformative columns **(2 points)**

Consider the following multiple alignment for five taxa `A,B,C,D,E`.

m columns

		1	2	3	4	5	6	7	8	9	10	11	12	
A	...	A	R	E	L	K	G	-	A	N	N	G	H	...
B	...	N	N	G	L	K	E	V	A	-	N	G	-	...
C	...	A	D	H	L	K	G	V	A	C	N	G	I	...
D	...	A	C	H	L	M	G	V	R	C	N	E	L	...
E	...	N	Q	H	I	M	E	V	R	C	N	G	-	...

(a) List all *uninformative* columns and their cost.

(b) List all *equivalent* columns.

Task 2 Maximum Parsimony Column-wise **(2 points)**

Find a binary most parsimonious tree for the given matrix under unit cost. We know that, for four taxa, these three topologies exist:

1. ((A,B),(C,D)); 2. ((A,C),(B,D)); 3. ((A,D),(B,C));

	1	2	3	4
A	d	b	x	z
B	d	b	x	y
C	c	a	e	z
D	c	a	e	y

Use the *Column-wise approach* to decide which topology allows for minimal cost. Write down all intermediate steps. If you can stop at some point explain why. (**Hint:** The order of the given topologies and states makes it a bit easier.)

Task 3 Greedy Sequencial Addition. **(2 points)**

Consider the *Greedy Sequencial Addition* heuristic (GSA) to approach a solution of the *maximum parsimony problem*. For given taxa $t_1, t_2, t_3, \ldots, t_n$, the algorithm works as described on page 35 in the lecture notes. See also Figure 6.3 on page 35.

(a) How many trees are considered before choosing one, if you add the i-th taxon?

(b) How many trees are considered in all steps $i = 4...n$? Show that this number increases quadratically with n. (Hint: Use the formula of the "little Gauß")

Task 4 Nearest Neighbor Interchange. **(2 points)**

Consider the given tree. List all trees that are *Neighbors* according to the *Nearest Neighbor Interchange* approach.