Week 9

This week we have been comparing trees obtained with ClonalTree and true trees by GCtree for **89 repertoires** (https://github.com/julibinho/ClonalTree/blob/main/Data/Simulations.zip).

We generated a new script in order to get the parameters that we will be using to measure the distance between the trees generated by the two different softwares. We are using:
Number of branches, 'PD - sum of length of all branches', 'avPD - average sum of length of all branches', 'Height 1', 'Height 2', 'Height 3', 'Depth 1', 'Depth 2', 'Depth 3', 'Size Tree - Profondeur d'arbre'.

We generated the mentioned parameters for our two trees (one from Clonal Tree and another from GCtree) form the 89 repertories, getting **Table 1** with 178 elements.

	Newick file	Number of branches	PD	avPD	H1	H2	H3	D1	D2	D3	SizeTree
0	30_1.clonalTree	12	13	1.08	2	1	0	0	1	2	3
1	30_1.GT.naive	12	13	1.08	2	1	0	0	1	2	3
2	30_2.clonalTree	12	14	1.17	2	0	0	1	2	2	4
3	30_2.GT.naive	12	14	1.17	2	0	0	1	2	2	4
4	30_3.clonalTree	5	6	1.20	1	0	0	0	1	1	2
...	...	...	...	...	...	...	...	...	...	...	...
173	200_8.GT.naive	95	112	1.18	4	2	0	0	1	1	5
174	200_9.clonalTree	64	77	1.20	3	2	1	0	1	1	4
175	200_9.GT.naive	64	77	1.20	3	2	1	0	1	1	4
176	200_10.clonalTree	88	104	1.18	4	3	2	1	1	3	6
177	200_10.GT.naive	88	107	1.22	4	3	2	1	1	2	6

178 rows × 11 columns

Table 1. Parameters of trees analyzed by Clonal Tree and GCtree

Using these parameters, we measured the Euclidean distance between each pair of trees coming from the same repertoire (for example, score for metrics for 30_1.clonalTree.nk and metrics for 30_1.GT.naive.nk). The distance was calculated using formula (1).

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_i - q_i)^2 + \cdots + (p_n - q_n)^2}.$$

(1)

From the 89 repertoires analyzed, we found that the distance between 51 of them was 0! This means that for 38 out of 98 repertoires ClonalTree and GCtree build trees with a rather different structure, and, consequently, we have different values for some of the metrics.

Finally, we analyzed the repertoires where we found differences in order to know where the most frequent differences were. We realized that the differences were mainly in the length of branches, and most specifically the ones in H1.

```
different Number of branches  =>  1

different PD  =>  30

different avPD  =>  30

different H1  =>  12

different H2  =>  6

different H3  =>  3

different D1  =>  1

different D2  =>  6

different D3  =>  9

different SizeTree  =>  18
```