

src/MetricsTree.py - a new file with our functions for metrics

Function **phyloDiversity(tree, filename)** that calculate:

- The number of branches
- Sum of all branch lengths (PD)
- Sum of all branch lengths divided by the number of clonotypes of the lineage (avPD)

input1 - ete3 tree

input2 - filename **.abRT.nk* (for example, *Examples/output/output30_1.abRT.nk*)

output - file **filename*MetricPD.txt* (*Examples/output/output30_1MetricPD.txt*)

* But we had problems with distances in Newick file created by ClonalTree, so, we made second function that use *.csv file with correct distances created by ClonalTree. *

Function **phyloDiversityCsv(filename)** that calculate:

- The number of branches
- Sum of all branch lengths (PD)
- Sum of all branch lengths divided by the number of clonotypes of the lineage (avPD)

input - filename **.abRT.nk* (for example, *Examples/output/output30_1.abRT.nk*)

output - file **filename*MetricPD.txt* (*Examples/output/output30_1MetricPD.txt*)

src/clonalTree.py changes

```
63     f.write(infoTree); f.close()
64
65     # functions for phyloDiversity parametres
66     # function phyloDiversity that use Newick sequence
67     # function phyloDiversityCsv that use csv file
68
69     #phyloDiversity(Tree, outputFile)
70     phyloDiversityCsv(outputFile)
71
72     print ('done')
73
```

We can make a key, for example, -m:

- if m == 0, file with the metrics will be created
- if m == 1, file with the metrics won't be created

Example of results: (For 30_1.fasta input sequence file)



output30_1MetricPD.txt

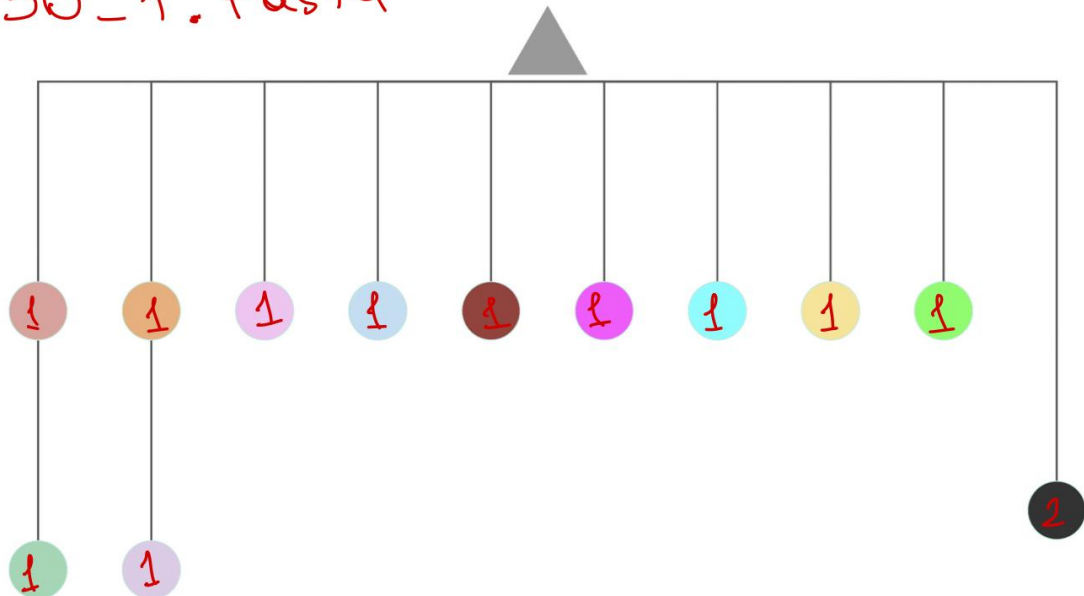
Number of branches : 12

Sum of all branches length (PD) : 13.0

Sum of all branch length divided by the number of clonotypes of the lineage (avPD) : 1.08

Visualisation of 30_1.fasta

30_1.fasta



1,2 - branch lengths

So for a given tree, let's calculate :

- The number of branches = 12
- Sum of all branch lengths (PD) = 13
- Sum of all branch lengths divided by the number of clonotypes of the lineage (avPD)

$$\frac{13}{12} = 1.08$$

?

Number of clonotypes : should we count naive cell or not?

If we shouldn't: number of clonotypes = number of branches

If we should: number of clonotypes = number of branches + 1