

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

Function **print_height_depth (file_fasta, tree, outputfile)** that calculate:

- The height for 3 nodes with the most high abundance in the tree
- The depth for 3 nodes with the most high abundance in the tree

input1 - fasta file with the sequences,

input2 - ete3 tree

input3 - filename - *.abRT.nk

output - file *filename*MetricPD.txt

Function **pathToNode(nodeName, nodeFin, tree)** that return:

- The path from nodeName to nodeFin

input1 - nodeName : name of first node of path,

input2 - nodeFin : name of last node of path (don't print in path)

input3 - ete3 tree

output - list of nodes in the path

src/clonalTree.py changes

```
#phyloDiversity(Tree, outputFile)
phyloDiversityCsv(outputFile)

# function that add height and depth for 3 most abundant nodes in Metrics.txt
print_height_depth (fastaFile, tree, outputFile)

print ('done')
```

Example of a file with our metrics:

```
dataset1_1_simplifieMetricPD.txt x
Number of branches : 16
Sum of all branches length (PD) : 39.0
Sum of all branch length divided by the number of clonotypes of the lineage (avPD) : 2.44
#=====#
Height
The most abundant clonotype - seq4 : height = 2
The second most abundant clonotype - seq12 : height = 1
The third most abundant clonotype - seq85 : height = 1
#=====#
Depth
The most abundant clonotype - seq4 : depth = 1
The second most abundant clonotype - seq12 : depth = 2
The third most abundant clonotype - seq85 : depth = 2
```

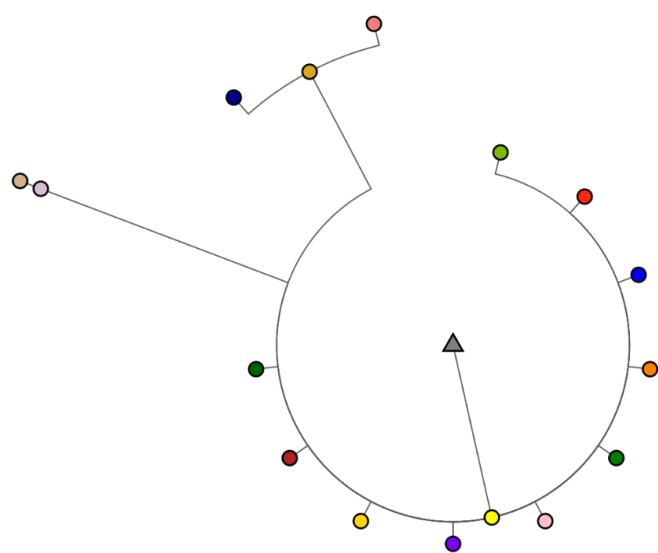
Sometimes we have an empty node. It writes like a 'none' node in the function **pathToRoot**. Should we count this node if it is on the path to count height and depth?

For example, in this dataset9_1:

- we have a rose node and we will have this 'none' node on the path to count depth ($D3 = 6$).
- we have a green node and we will have this 'none' node on the path to count height ($H2 = 5$).

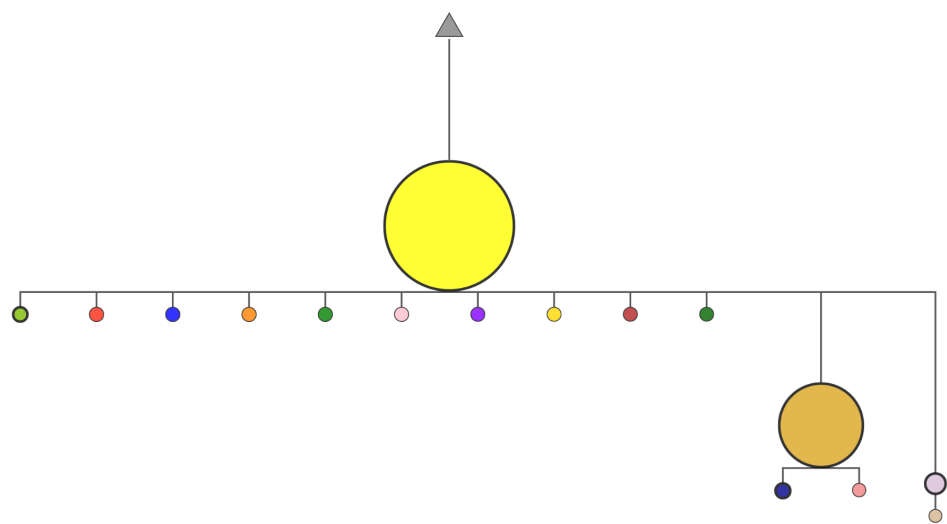
	Categorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage	H1 - height of the most abundant	H2 - height of the second most abundant	H3 - height of the third most abundant	D1 - depth of the most abundant	D2 - depth of the second most abundant	D3 - depth of the third most abundant
dataset1_1_simplifie	1	16	39	2,44	2	1	1	1	2	2
dataset2_1_simplifie	1	17	31	1,82	2	1	0	2	3	3
dataset3_1_simplifie	1	17	53	3,12	2	1	0	3	4	4
dataset4_1_simplifie	1	16	63	3,94	2	1	0	3	4	4
dataset5_1_simplifie	1	17	46	2,71	4	1	0	2	5	3
dataset6_1_simplifie	2	18	38	2,11	1	2	0	3	2	4
dataset7_1_simplifie	2	17	33	1,94	1	1	0	2	1	3
dataset8_1_simplifie	3	22	34	1,55	1	0	3	3	1	3
dataset9_1_simplifie	3	19	74	3,89	1	5	0	3	2	6
dataset10_1_simplifie	3	17	49	2,88	1	4	2	5	2	4
dataset11_2_simplifie	4	28	170	6,07	1	1	1	3	2	2
dataset12_1_simplifie	4	18	47	2,61	1	1	0	2	2	2
dataset13_1_simplifie	4	22	139	6,32	1	3	1	5	2	4

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset1_1_simplifie	1	16	39	2,44



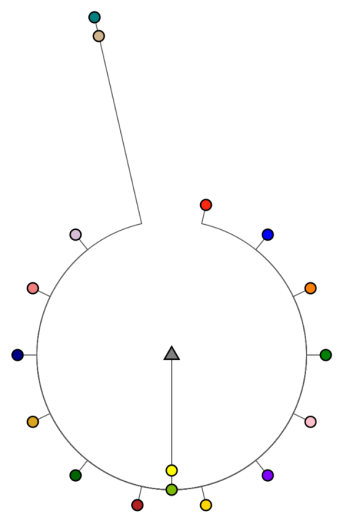
dataset1_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset1_1_simplifie	2	1	1	1	2	2



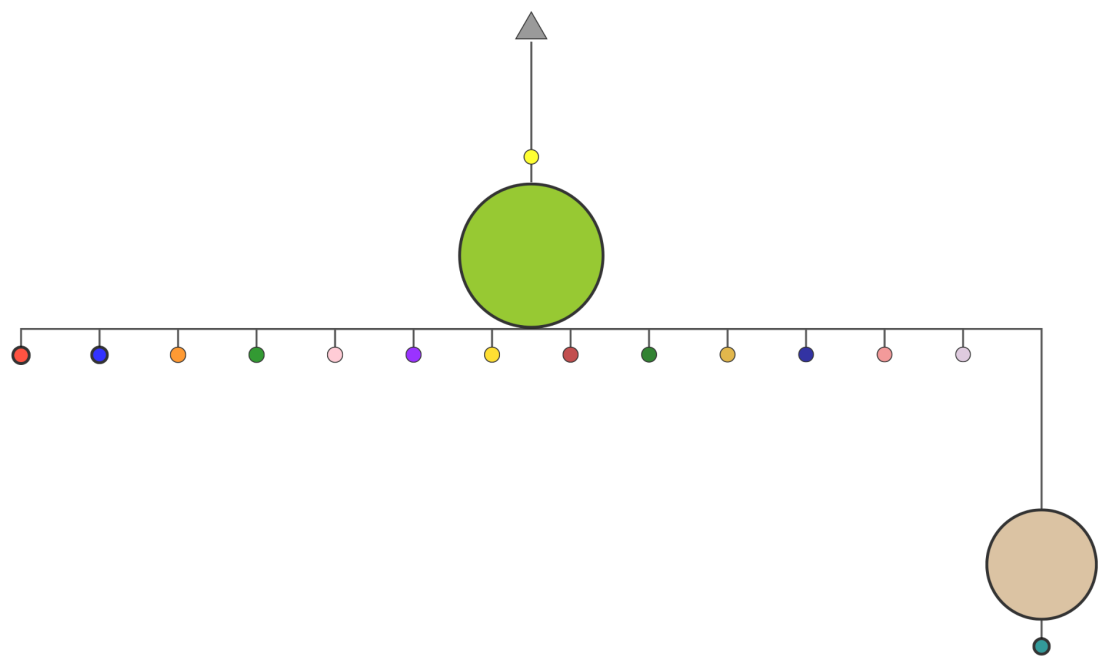
dataset1_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset2_1_simplifie	1	17	31	1,82



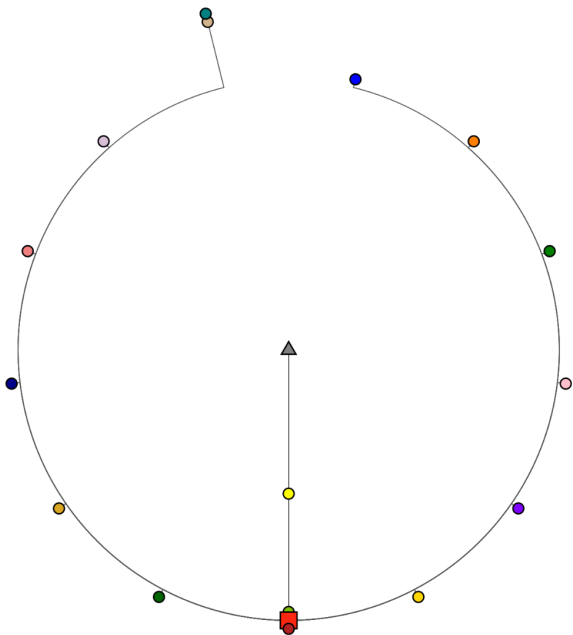
dataset2_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset2_1_simplifie	2	1	0	2	3	3



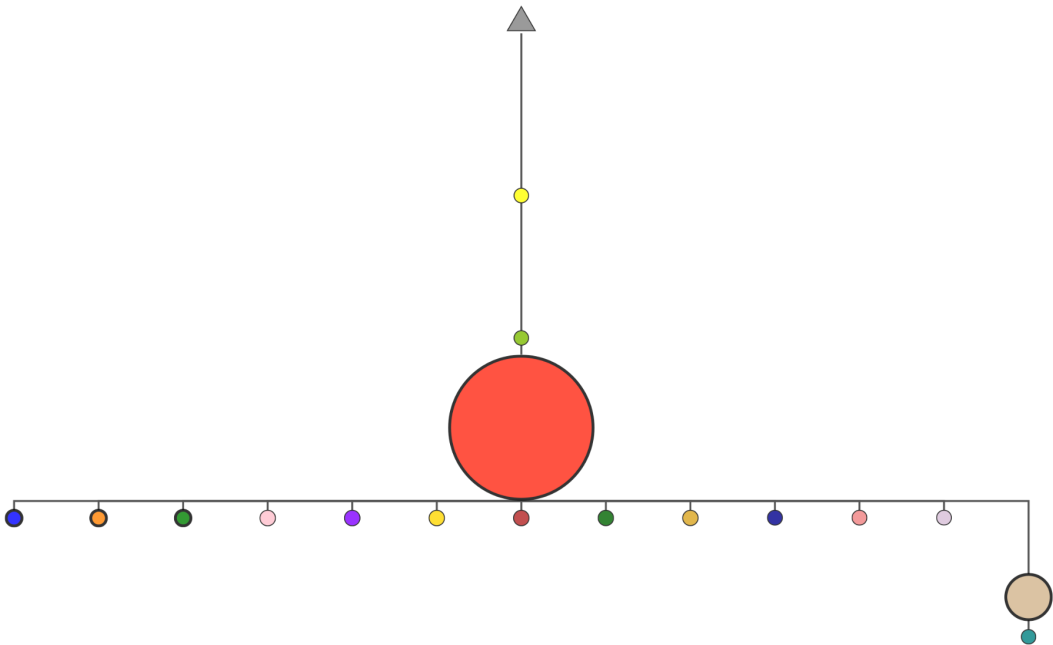
dataset2_1 simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset3_1_simplifie	1	17	53	3,12



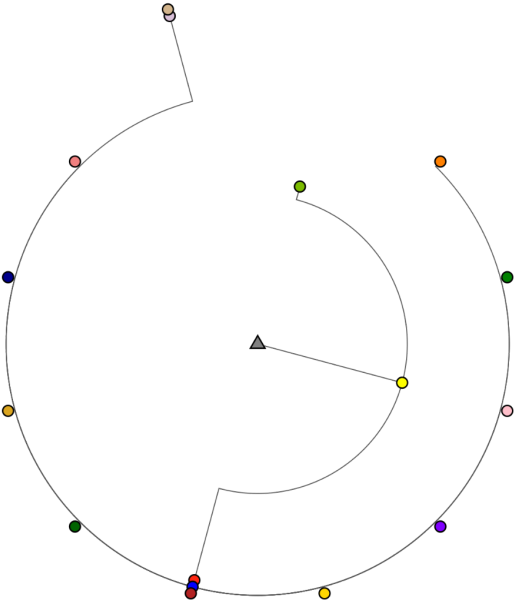
dataset3_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset3_1_simplifie	2	1	0	3	4	4



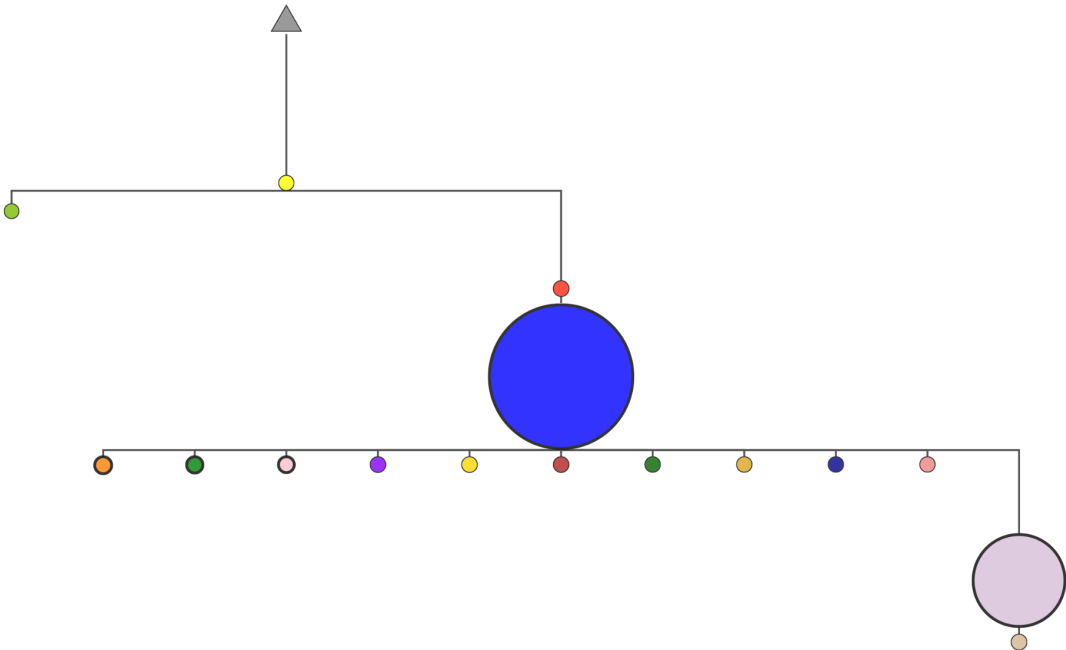
dataset3_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset4_1_simplifie	1	16	63	3,94



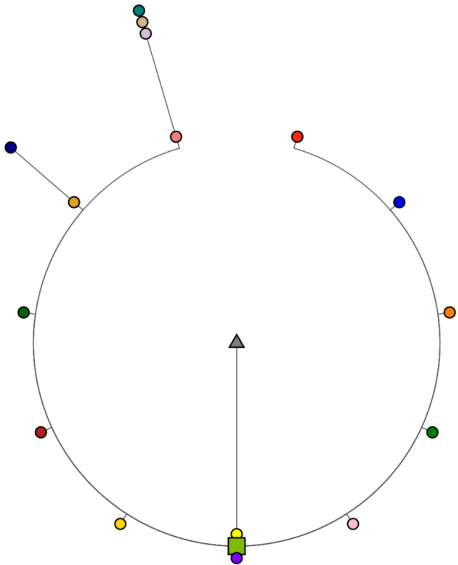
dataset4_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset4_1_simplifie	2	1	0	3	4	4



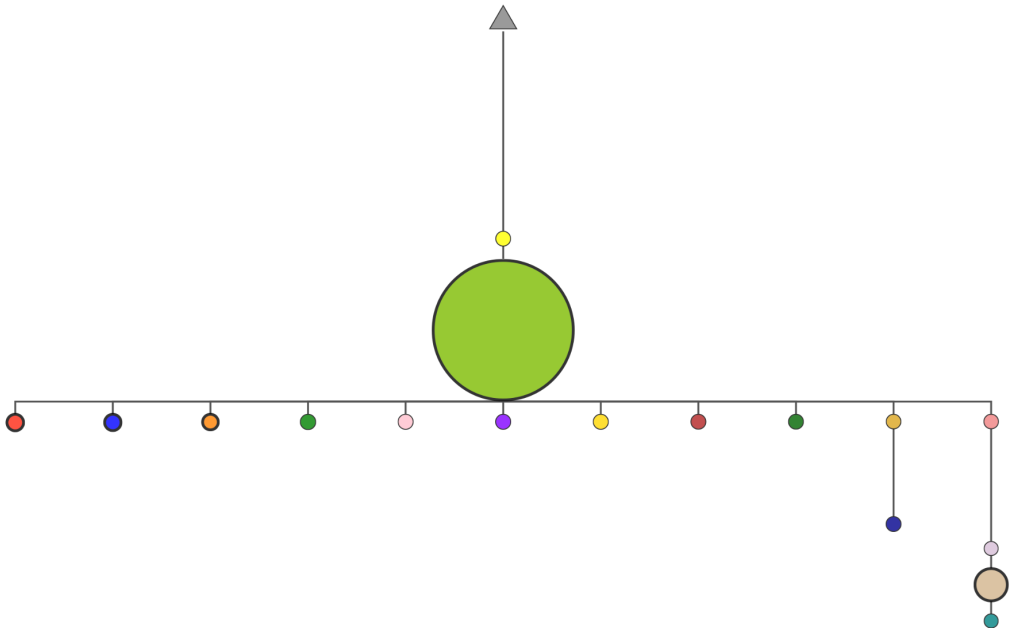
dataset4_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset5_1_simplifie	1	17	46	2,71



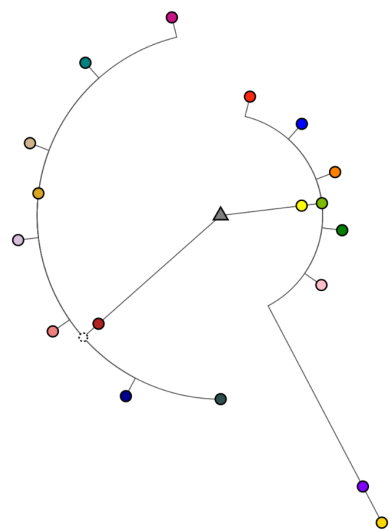
dataset5_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset5_1_simplifie	4	1	0	2	5	3



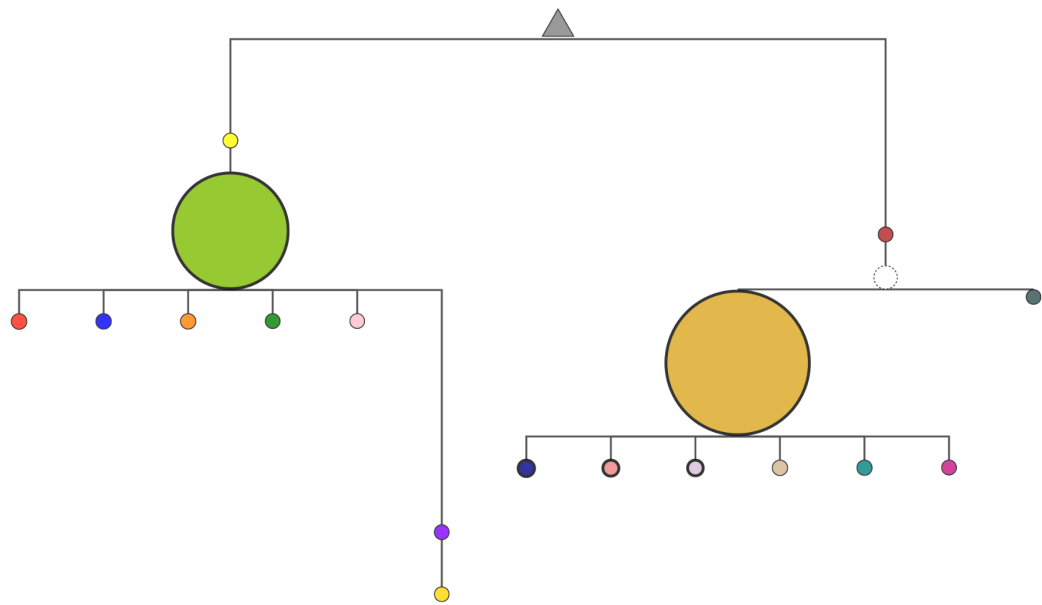
dataset5_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset6_1_simplifie	2	18	38	2,11



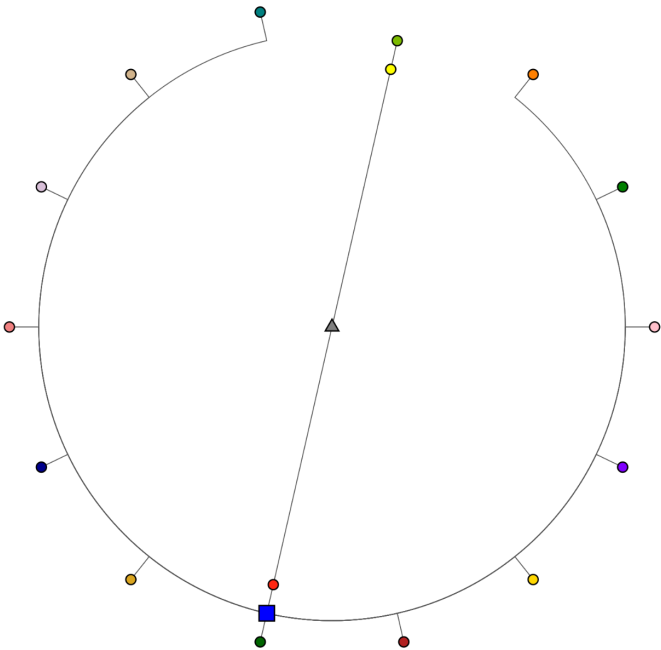
dataset6_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset6_1_simplifie	1	2	0	3	2	4



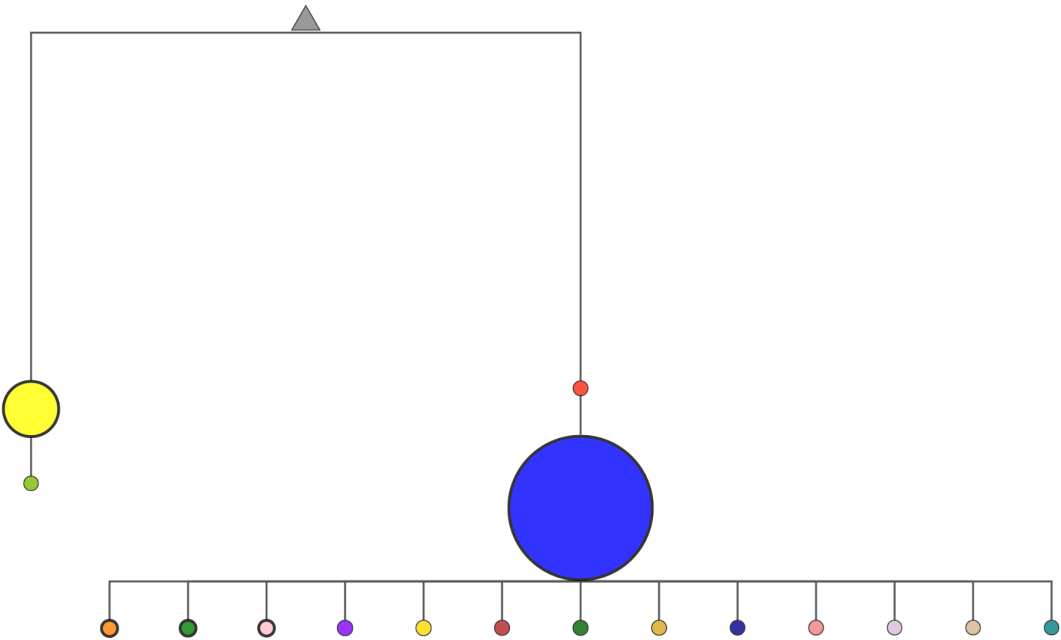
dataset6_1 simplifie Elbow tree

	Categorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset7_1_simplifie	2	17	33	1,94



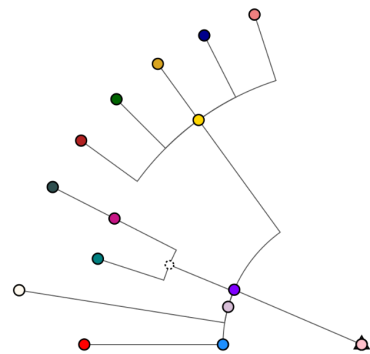
dataset7_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset7_1_simplifie	1	1	0	2	1	3



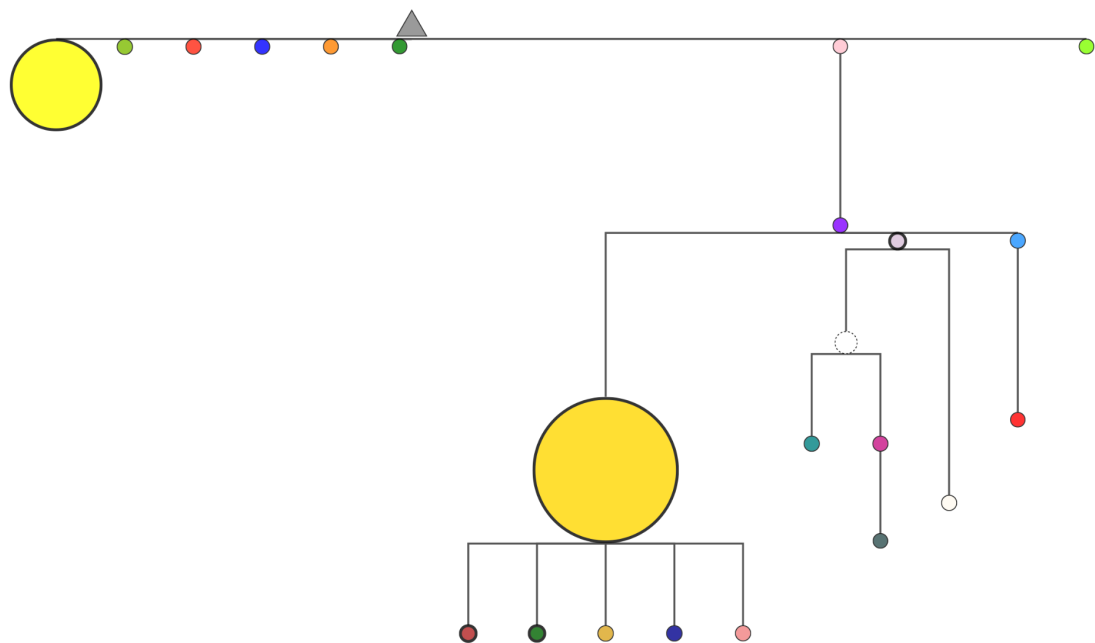
dataset7_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset8_1_simplifie	3	22	34	1,55



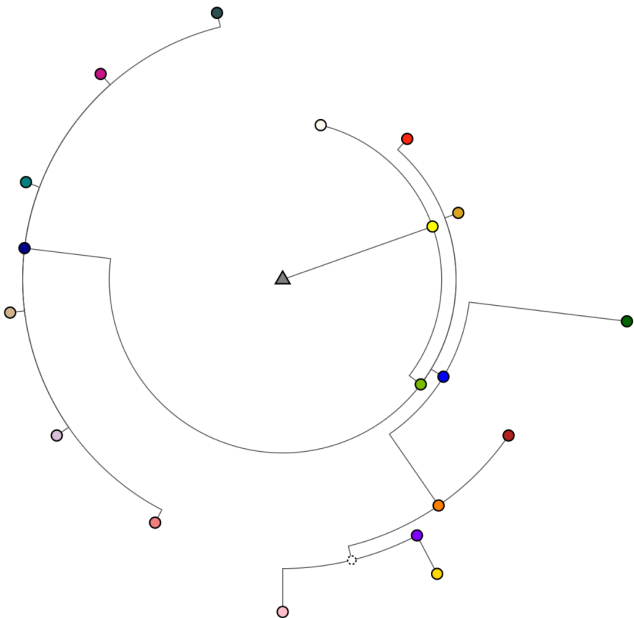
dataset8_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset8_1_simplifie	1	0	3	3	1	3



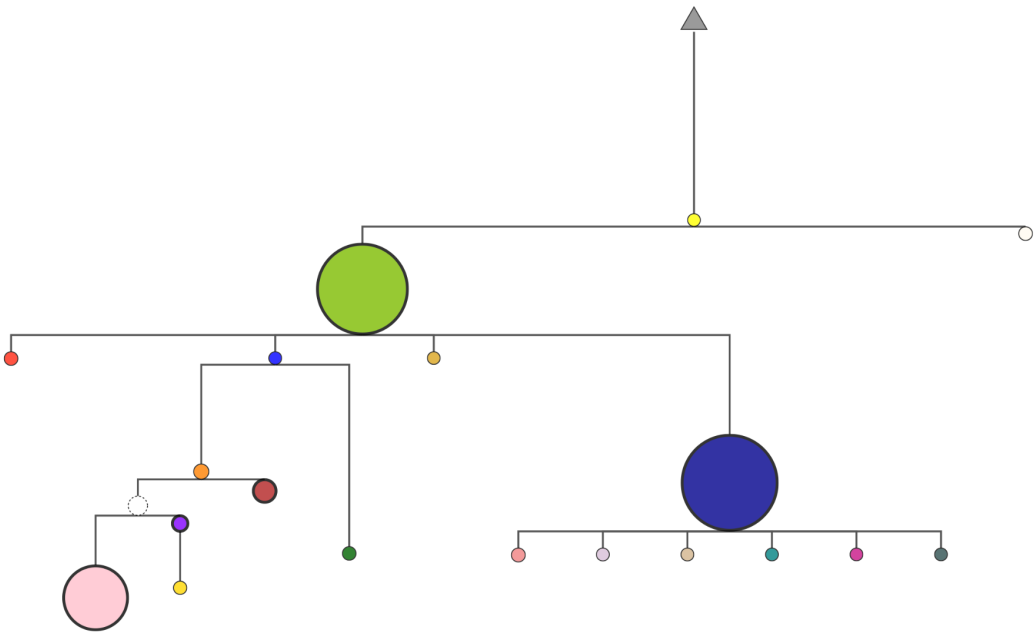
dataset8_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset9_1_simplifie	3	19	74	3,89



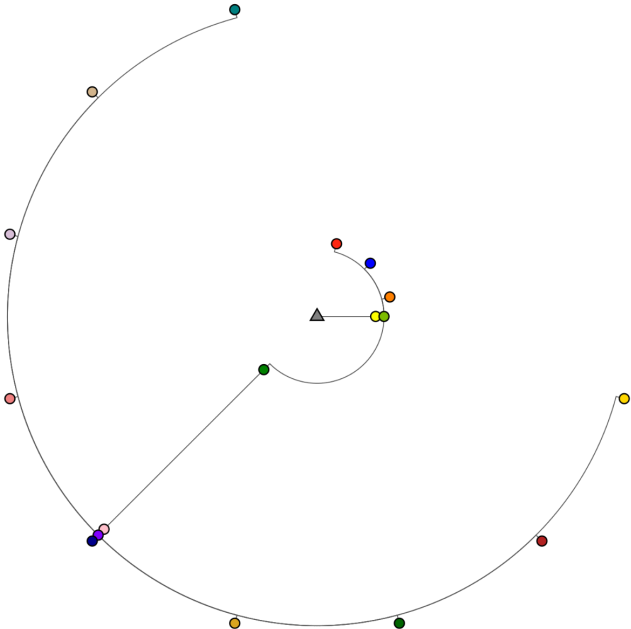
dataset9_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset9_1_simplifie	1	5	0	3	2	6



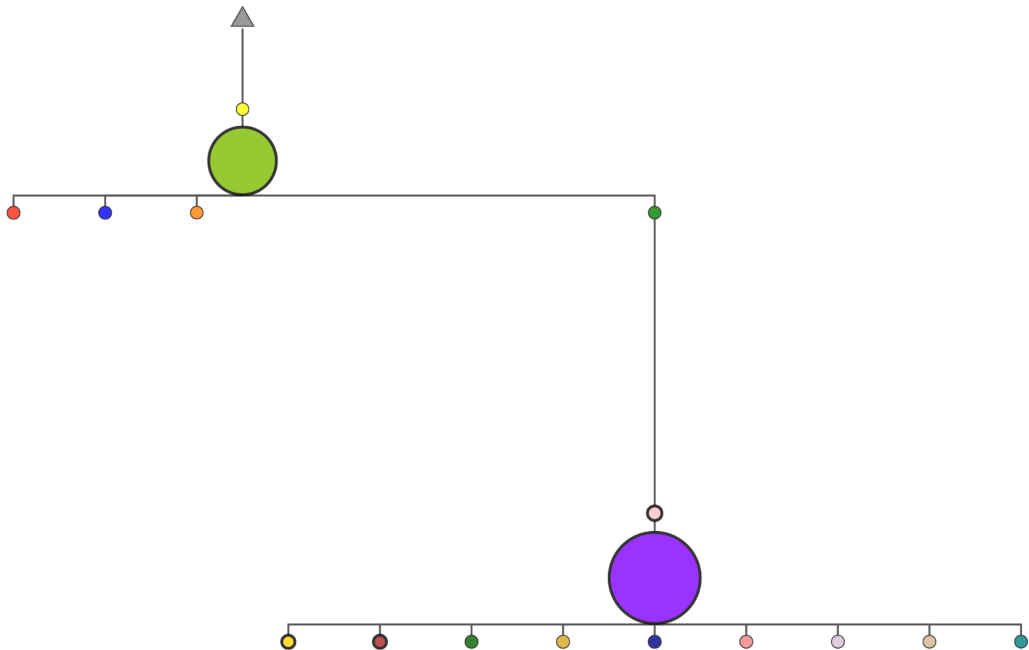
dataset9_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset10_1_simplifie	3	17	49	2,88



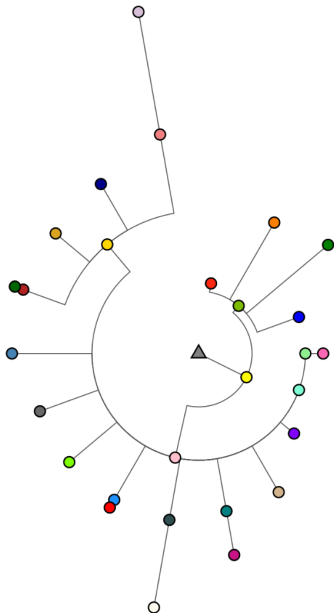
dataset10_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset10_1_simplifie	1	4	2	5	2	4



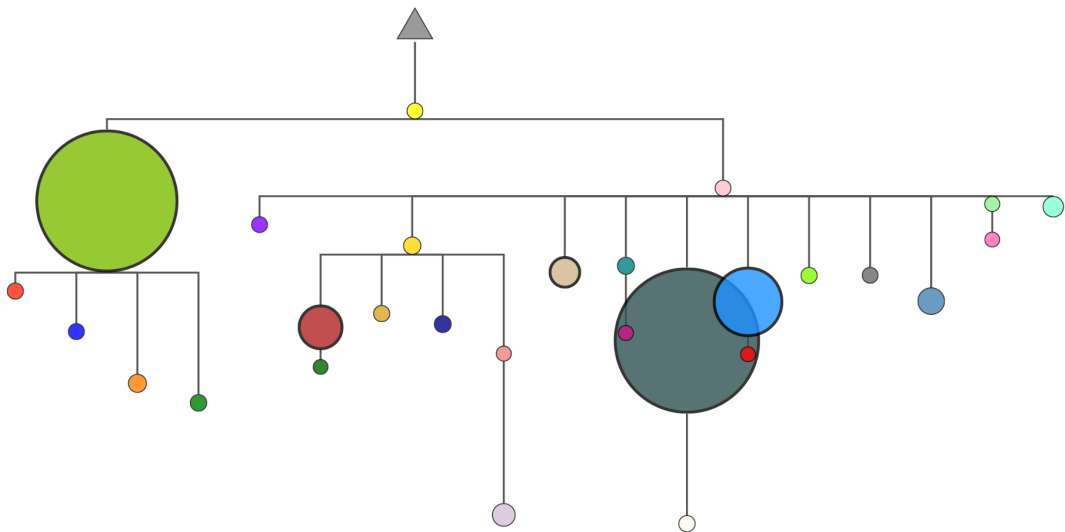
dataset10_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset11_2_simplifie	4	28	170	6,07



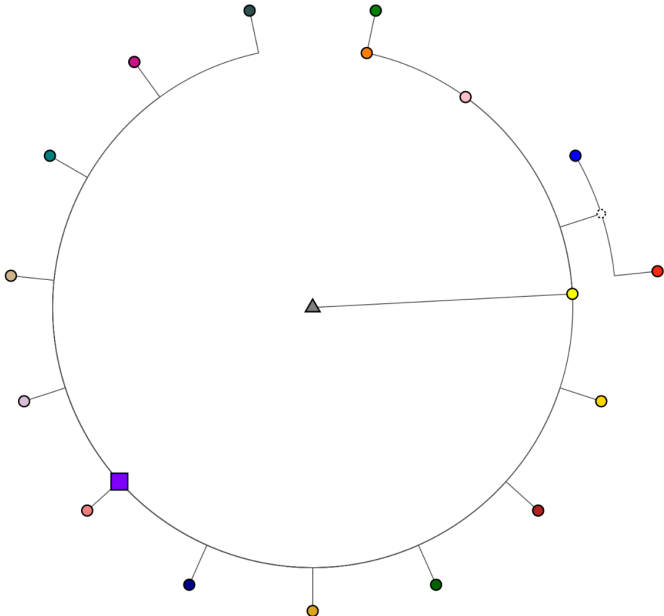
dataset11_2_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset11_2_simplifie	1	1	1	3	2	2



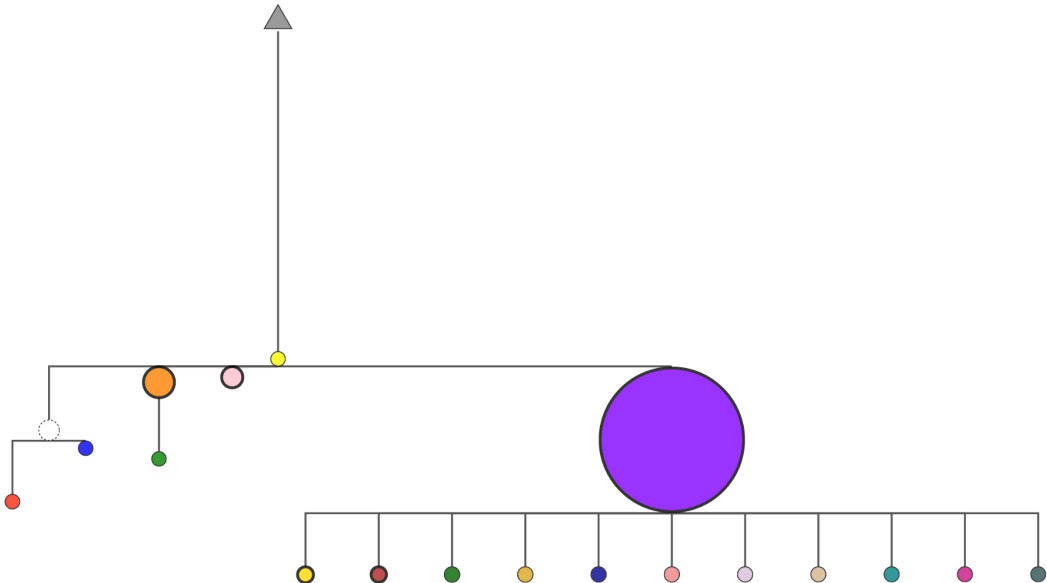
dataset11_2_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset12_1_simplifie	4	18	47	2,61



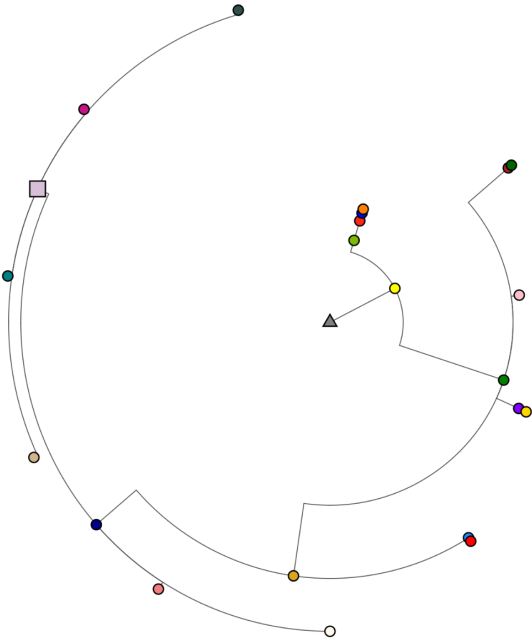
dataset12_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset12_1_simplifie	1	1	0	2	2	2



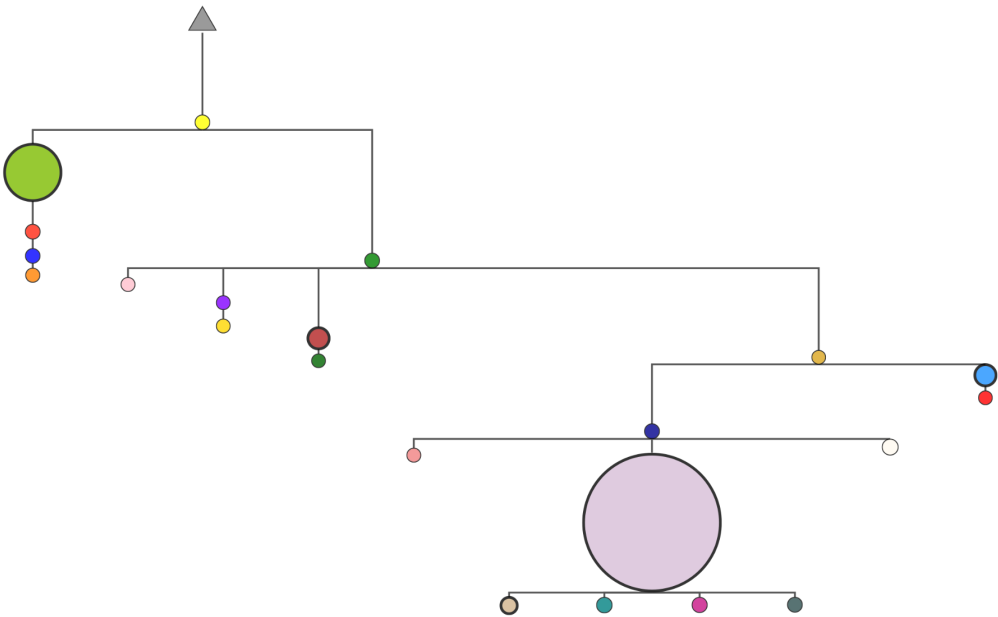
dataset12_1_simplifie Elbow tree

	Catégorie	Number of branches	(PD) - Sum of all branches length	(avPD) - Sum of all branch length divided by the number of clonotypes of the lineage
dataset13_1_simplifie	4	22	139	6,32



dataset13_1_simplifie Circle tree

	H1	H2	H3	D1	D2	D3
dataset13_1_simplifie	1	3	1	5	2	4



dataset13_1_simplifie Elbow tree