

CellPhy - Mutation mapping

Alexey Kovlov, João M Alves*, Alexandros Stamatakis & David Posada

June 2020

1 Introductory Note

In the following supporting document, we will go through all steps required to generate the results from **CellPhy**'s mutation mapping function. For reproducibility purposes, this step-by-step tutorial contains all information needed to produce the figure 6.A presented in the main manuscript.

2 Mapping mutations onto a phylogenetic tree

Cancer genomics studies are, for the most part, interested in understanding when “*driver*” mutations appeared in the malignant cell population. Synonymous mutations are generally thought to be functionally silent and evolutionarily neutral, so our focus here will be to solely map the non-synonymous mutations present in our SNV set. On this basis, we will take as input a thinned VCF only carrying the non-synonymous mutations together with the best tree and model estimates from our previous CellPhy run (tree search + bootstrap).

```
$ MODEL=CRC24.raxml.bestModel
$ TREE=CRC24.raxml.bestTree

$ raxml-ng --mutmap \
  --msa CRC24.NS.vcf \
  --model $MODEL --tree $TREE --opt-branches off \
  --prefix CRC24.NS_Mapped --threads 1
```

3 Visualizing the results

Once it's done, CellPhy should output 2 distinct files:.

- **(A)** *CRC24.NS_Mapped.mutationMapTree*
→ Newick tree file with indexed branches
- **(B)** *CRC24.NS_Mapped.mutationMapList*
→ Text file with the number and the list of mutations per branch

cellphy-mutationmapping.sh can now be used to plot the mutations onto the inferred phylogenetic tree. If you run it without any parameters, it will show a help message:

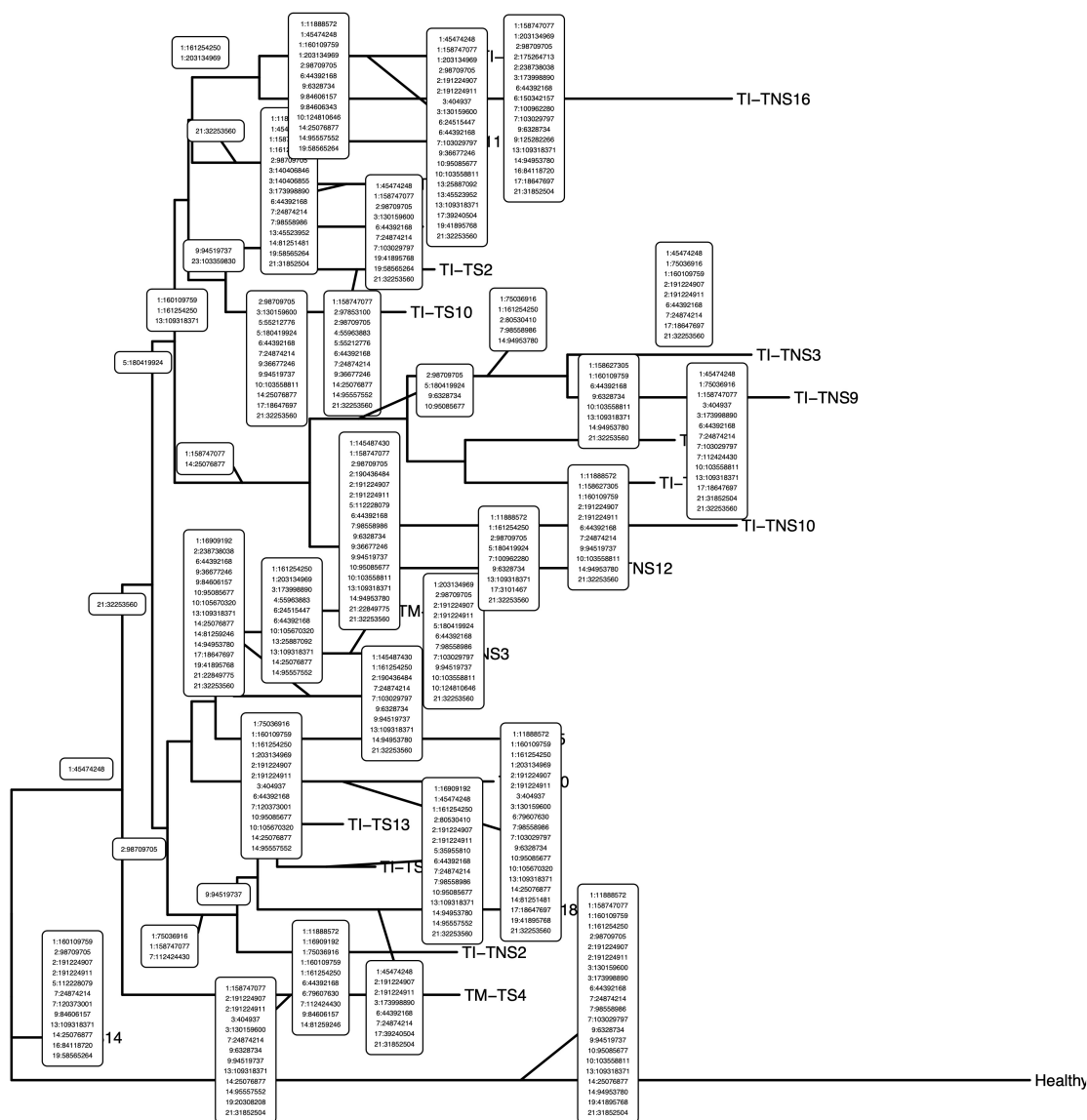
```
$ ./cellphy-mutationmapping.sh
Usage: ./cellphy-mutationmapping.sh raxml.mutationMapTree raxml.mutationMapList Output_prefix [geneIDs]
Created by: Alexey Kovlov, Joao M Alves, Alexandros Stamatakis & David Posada - 16 June 2020
*Required files:
  -Tree
  -Mutation List
  -Output Prefix
```

*Optional:
-Gene IDs (Tab-delimited)

Now let's run it again but this time with the required parameters:

```
$ ./cellphy-mutationmapping.sh CRC24.NS_Mapped.mutationMapTree CRC24.NS_Mapped.mutationMapList CRC24.Mapped
Generating tree plot...
Done!
```

If everything went as expected, you should have generated the following figure, in PDF format (*CRC24.Mapped.Tree_mapped.pdf*), where the mutations are mapped onto the tree branches:



If you are interested in plotting the gene names instead, you can provide a tab-delimited file linking the genomic position to its gene ID:

```
$ head -n5 CRC24.NS.GeneIDs
#chr:pos gene
1:11888572 CLCN6
1:16909192 NBP1
1:45474248 HECTD3
1:75036916 C1orf173
```

Afterwards, we can run *cellphy-mutationmapping.sh* again, but changing the output prefix so that you don't overwrite the previous results:

```
$ ./cellphy-mutationmapping.sh CRC24.NS_Mapped.mutationMapTree CRC24.NS_Mapped.mutationMapList CRC24.GeneID.Mapped CRC24.NS.GeneIDs
Converting positions to GeneID...
Done!
Generating tree plot...
Done!
```

You will notice that our tree now has the gene names displayed, instead of the genomic positions:

