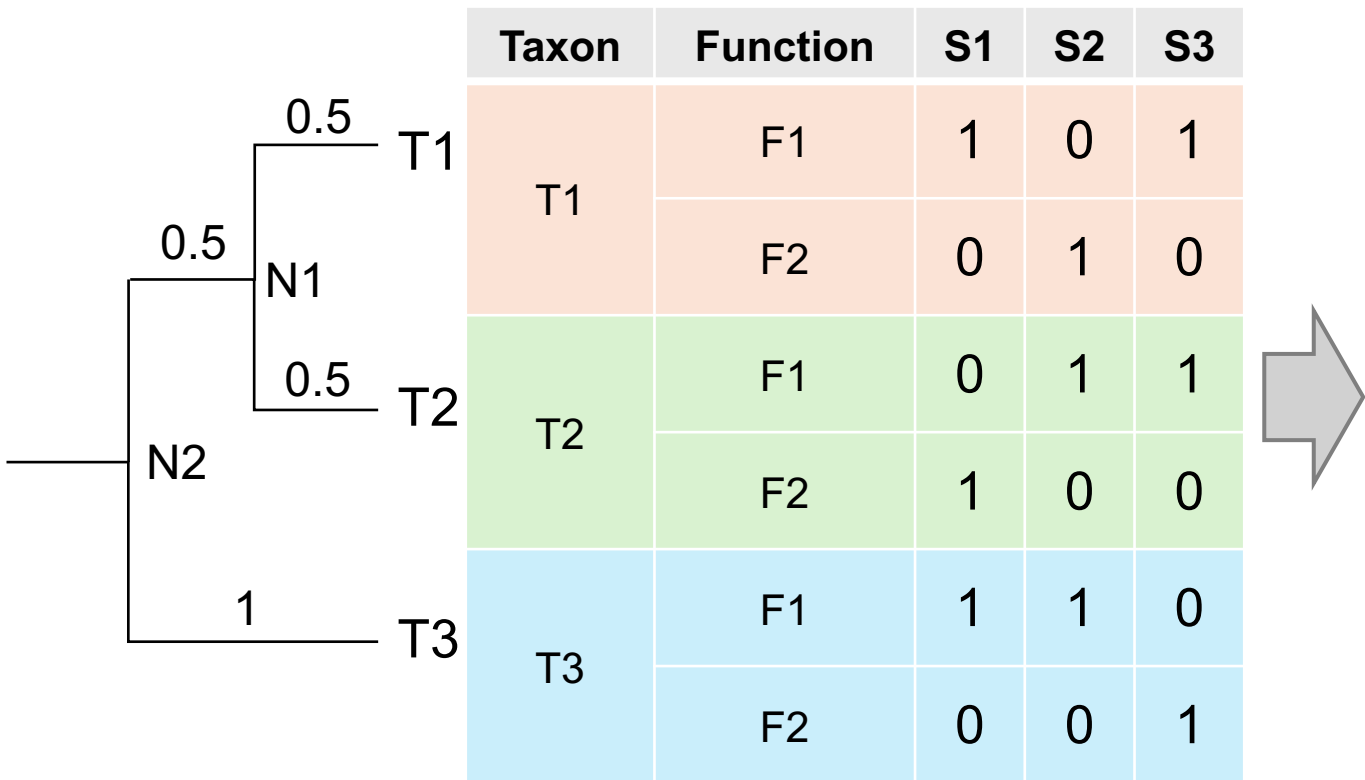
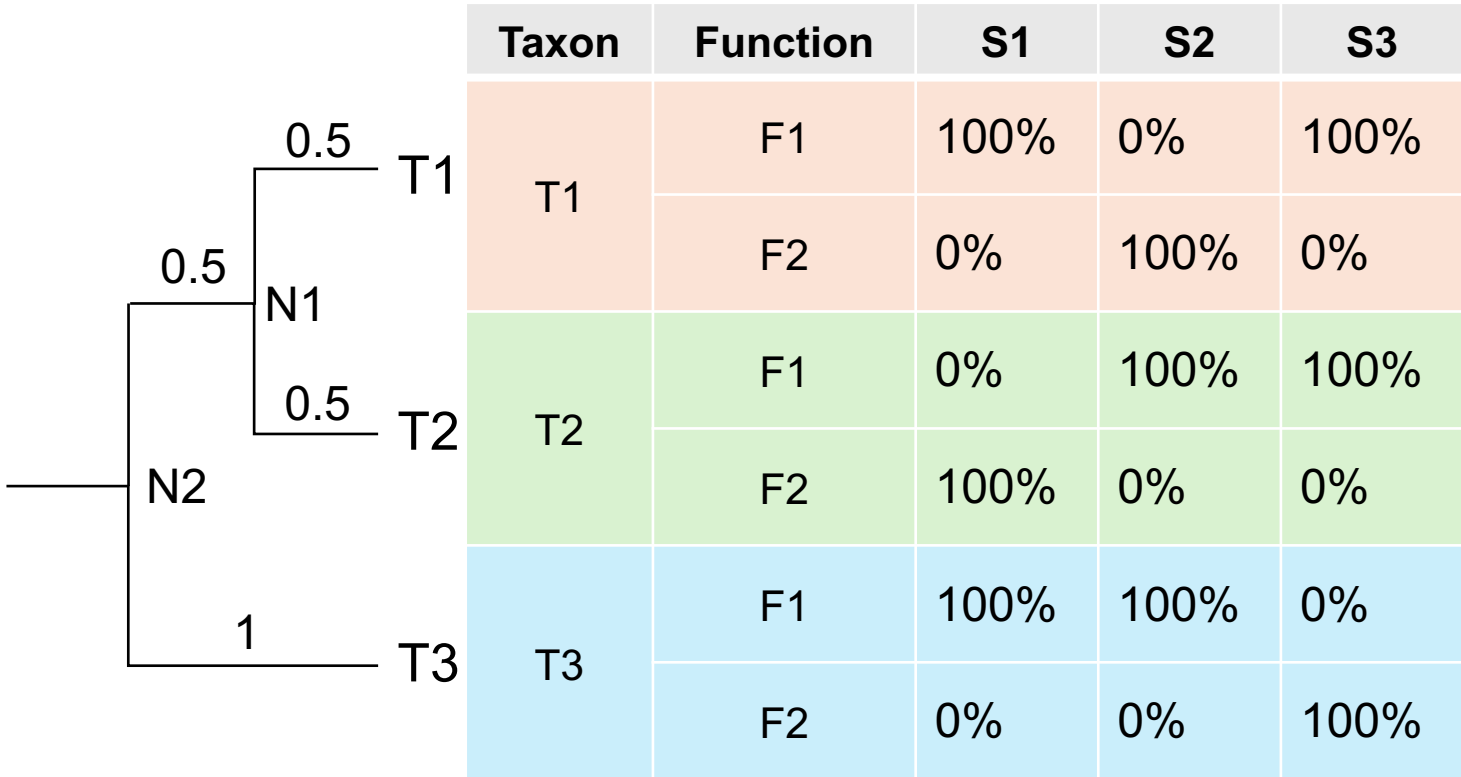


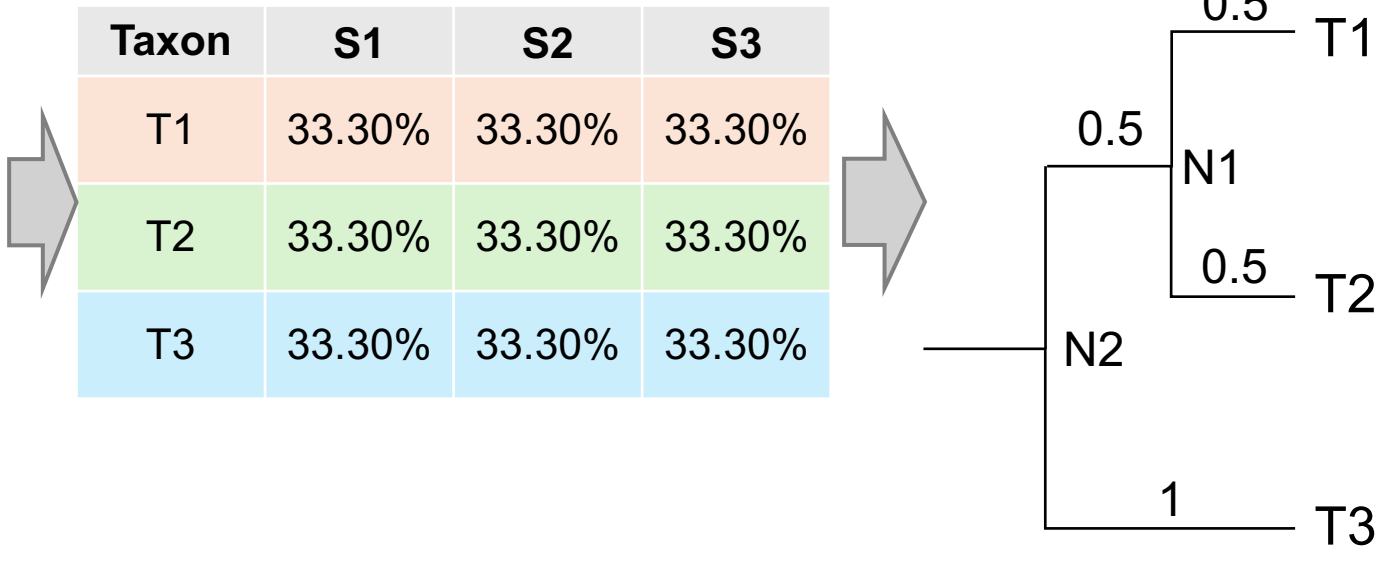
**Step 1.**  
Obtain phylogeny-informed taxon-function data,  
containing tree structure, branch lengths,  
functional abundance of each taxa.



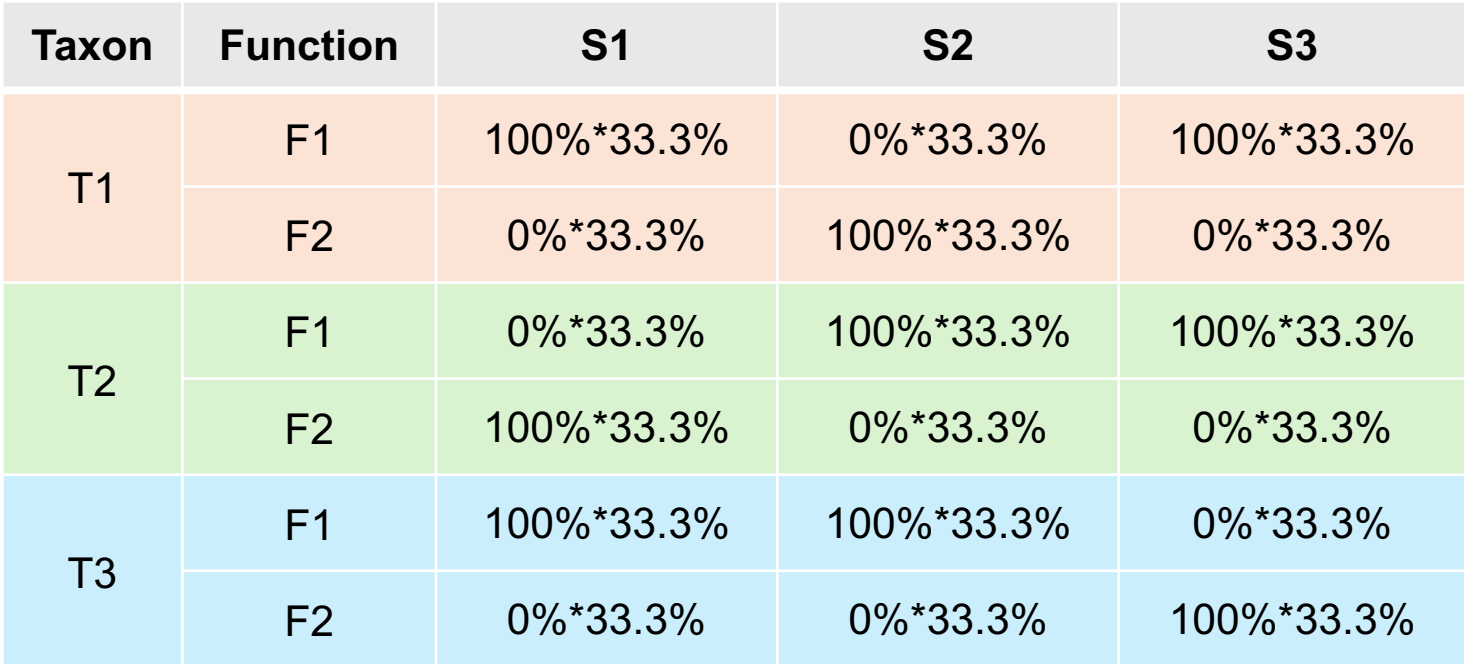
**Step 2.**  
Calculate relative abundance of functions within each taxa.



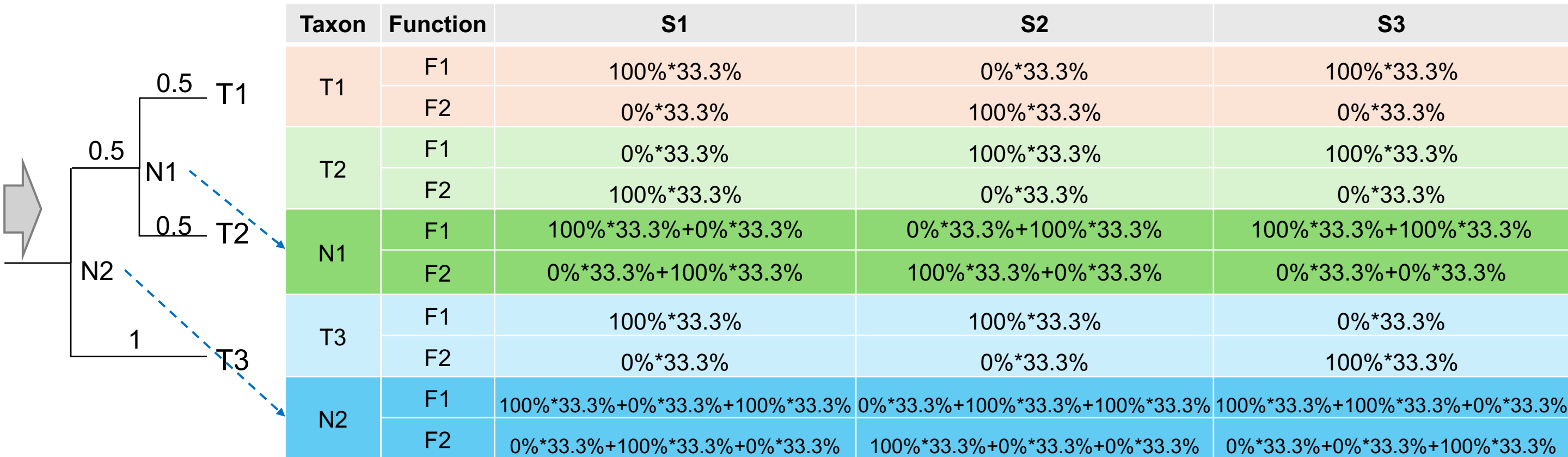
**Step 3.**  
Calculate relative abundance of taxa by  
taxon-specific protein biomass contributions.



**Step 4.**  
Further weigh the relative functional abundance by their  
corresponding relative taxonomic abundance.



**Step 5.**  
The table is expanded to represent all nodes up to the root of the phylogeny through summing up each node.



**Step 6.**  
Calculate relative abundance of functions within each tree node.

Taxon	Function	S1	S2	S3
T1	F1	100.0%	0.0%	100.0%
	F2	0.0%	100.0%	0.0%
T2	F1	0.0%	100.0%	100.0%
	F2	100.0%	0.0%	0.0%
N1	F1	50.0%	50.0%	100.0%
	F2	50.0%	50.0%	0.0%
T3	F1	100.0%	100.0%	0.0%
	F2	0.0%	0.0%	100.0%
N2	F1	66.6%	66.6%	66.6%
	F2	33.3%	33.3%	33.3%

**Step 7.**  
Similarly, convert the taxonomic table into an expanded table.

Taxa	S1	S2	S3
T1	33.3%	33.3%	33.3%
T2	33.3%	33.3%	33.3%
N1	66.7%	66.7%	66.7%
T3	33.3%	33.3%	33.3%
N2	100.0%	100.0%	100.0%

**Step 8.**  
Next calculate functional distances between each sample pair, i.e.  $d_{(s1, s2)}$ ,  $d_{(s1, s3)}$ ,  $d_{(s2, s3)}$ , at each branch node, according to weighted Jaccard distance:  $d_{ab} = 1 - \frac{\sum_i^{\phi} \min(F_{ia}, F_{ib})}{\sum_i^{\phi} \max(F_{ia}, F_{ib})}$ , where  $a$  and  $b$  are the two samples,  $\phi$  denotes total number of functions,  $F_{ia}$  and  $F_{ib}$  represent normalized functional abundance of the  $i$ th function in samples  $a$  and  $b$ , respectively.

$d_{(s1, s2)} = 1 - \frac{0 + 0}{100 + 100} = 1$	$d_{(s1, s3)} = 1 - \frac{100 + 0}{100 + 0} = 0$	$d_{(s2, s3)} = 1 - \frac{0 + 0}{100 + 100} = 1$
$d_{(s1, s2)} = 1 - \frac{50 + 50}{50 + 50} = 0$	$d_{(s1, s3)} = 1 - \frac{50 + 0}{100 + 50} = 0.667$	$d_{(s2, s3)} = 1 - \frac{50 + 0}{100 + 50} = 0.667$
$d_{(s1, s2)} = 1 - \frac{0 + 0}{100 + 100} = 1$	$d_{(s1, s3)} = 1 - \frac{0 + 0}{100 + 100} = 1$	$d_{(s2, s3)} = 1 - \frac{100 + 0}{100 + 0} = 0$
$d_{(s1, s2)} = 1 - \frac{66.6 + 33.3}{66.6 + 33.3} = 0$	$d_{(s1, s3)} = 1 - \frac{66.6 + 33.3}{66.6 + 33.3} = 0$	$d_{(s2, s3)} = 1 - \frac{66.6 + 33.3}{66.6 + 33.3} = 0$
$d_{(s1, s2)} = 1 - \frac{100 + 0}{100 + 0} = 0$	$d_{(s1, s3)} = 1 - \frac{0 + 0}{100 + 100} = 1$	$d_{(s2, s3)} = 1 - \frac{0 + 0}{100 + 100} = 1$

**Step 9.**  
Finally, weigh  $d_{(s1, s2)}$  (as well as  $d_{(s1, s3)}$ ,  $d_{(s2, s3)}$ ) at each branch node by branch length and relative protein abundances in both samples, to obtain  $PiF_{(s1, s2)}$  (as well as  $PiF_{(s1, s3)}$ ,  $PiF_{(s2, s3)}$ ).

Taxa	$d_{(s1, s2)}$	Branch length	Sample 1, abundance	Sample 2, abundance	Weighted distance
T1	1	0.5	33.3%	33.3%	0.0554
T2	1	0.5	33.3%	33.3%	0.0554
N1	0	0.5	66.7%	66.7%	0
T3	0	1	33.3%	33.3%	0
N2	0	1	100.0%	100.0%	0

**Sums up to**  
 $PiF_{(s1, s2)} = 0.111$

**Similarly, we have:**  
 $PiF_{(s1, s3)} = 0.3147$   
 $PiF_{(s2, s3)} = 0.3147$