

Assignment 4.1

Comparative genomics

Stefano Pellegrini (mlq211)

November 6, 2020

1 Ancestral state and mode of evolution

1.1 Which is the ancestral state of the OBP gene family size in the phylogeny?

I run CAFE using as input the OBP gene counts and the phylogenetic tree to estimate the ancestral state reconstruction of the OBP gene family. The resulting output tree (**Base_asr.tre**) with the gene family size across all the phylogeny is the following:

```
((((( ((( (dmel<5>_52:5,(dsec<1>_51:1,dsim<0>_53:1)<4>_52:4)<7>_52:8,(dyak
<3>_55:10,dere<2>_50:10)<6>_52:3)<11>_52:31,dana<10>_50:44)<15>_50:11,(
dpse<9>_45:1,dper<8>_45:1)<14>_45:54)<19>_49:7,dwil<18>*_62:62)<21>_49
:1,((dmoj<13>_43:31,dvir<12>_41:31)<17>_44:12,dgri<16>_46:43)<20>_46
:20)<23>_49:187,agam<22>*_83:250)<25>_44:10,bmor<24>_46:260)<27>_43:20,
tcas<26>_49:280)<31>_40:20,amel<30>_21:300)<33>_36:50,(phum<29>*_5:260,
apis<28>_18:260)<32>*_21:90)<34>_30;
```

Each “_n” represents the ancestral state of each node, it follows the relative node ID embedded in the “<>” characters. The ancestral states across all the phylogeny can be better visualized pasting the tree in the website icytree.org (internal nodes text labels should be activated). The result is shown in Figure 1.

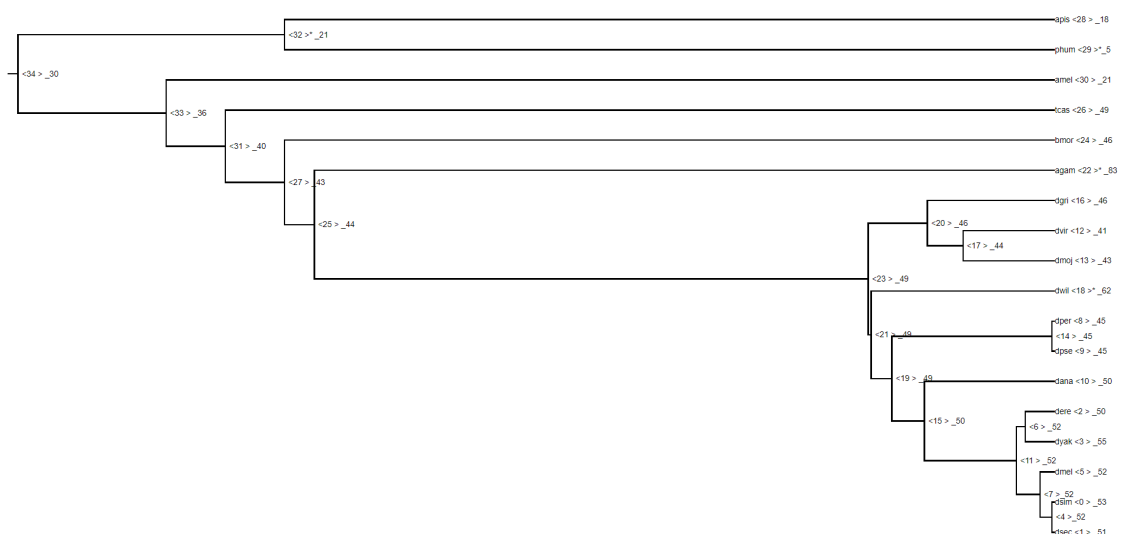


Figure 1: Output tree visualization by icytree.org

1.2 Which is the general mode of evolution of the OBP gene family across Hexapoda?

The file `Base_family_results.txt` contains the result for testing the hypothesis that the OBP gene family has undergone expansion or contraction in the surveyed species. Since the p-value for this hypothesis is 0.001 (the null hypothesis is that there is no change and therefore that λ equal 0) we can state that the OBP gene family size has undergone significative changes (expansion or contraction) across the different branches of the phylogeny. The overall entity of this change in the 18 surveyed *Hexapoda* species is described by the birth-and-death rate parameter λ , which is equal to 0.0033.

1.3 Has it undergone different dynamics across the surveyed species?

Yes, it has undergone different dynamics across the surveyed species. Looking at the `Base_change.tab` and `Base_branch_probabilities.tab` files, we can see in which species it has undergone significantly expansion or contraction. We can observe (Table 1) that in 3 species (*dwil*, *agam*, and *phum*) out of 18, the OBP gene family has undergone significant changes (posterior probability threshold of 0.05). In particular, it has undergone significant expansion in *dwil* and *agam* and significant contraction in *phum* species.

Table 1: Species in which OBP gene family has undergone significant contraction or expansion.

Species	Change	Posterior probability
dwil	13	0.005
agam	39	8e-05
phum	-16	0.001

2 Birth-and-death under two-model lambda

2.1 Are there differences in the OBP dynamics between *Drosophila* and the other Hexapoda species surveyed?

In order to answer this question I run CAFE several times and I compared the likelihood obtained by the one-lambda model to the one obtained by the two-lambda model. Since the two-lambda model obtained a better likelihood (approximately -80 for the one-lambda model and -79 for the two-lambda one), I can state that there are differences in the OBP dynamics between *Drosophila* and the other *Hexapoda* species surveyed.

2.2 Is the birth-and-death process constant across the phylogeny?

Since the two-lambda model, which estimated two different death-and-birth rates for *Drosophila* and the other *Hexapoda*, shown a better fit to the data than the one-lambda model, we can state that the birth-and-death process is not constant across the phylogeny.

3 Functional constraints

3.1 Which is the dN/dS ratio in the three duplicated OBP genes?

The dN/dS values for Obp56f, Obp56g, and Obp56h are respectively 0.338, 0.183, 0.546.

3.2 Are there different functional constraints among copies? Which is the most likely reason of this observation?

Observing the dN/dS values it is possible to see that the OBP duplicated genes undergone negative purifying selection and they have different functional constraints. In fact, when duplication occurred the OBP gene copies had the same dN/dS and the same functional constraints, but since the duplicated genes perform the same function, the negative selection could have been relaxed in a different way between the copies (Obp56g having the strongest functional constraint and Obp56h the weakest one), leading to a different dN/dS and different functional constraints.

3.3 Is there any specific site under positive selection?

I used MEME to infer sites that are evolving under positive selection. My result (p-value threshold of 0.1) is that one specific site (position 72) is under positive selection.