The odorant binding proteins (OBPs) are globular in nature and specifically transport semiochemicals through sensillum lymph to odorant receptors (ORs) that ultimately send electrical signals to the insect brain

Assignment 4 – Part 2: Comparative genomics

The main objective of this exercise is to reproduce by yourself the comparative genomics analyses learned during these sessions.

The datasets needed for the assignment can be found at: https://drive.google.com/drive/folders/1AFOuRtZS5uJ0rpQHfntsKudjuIEtw5L8?usp=sharing or in the student server (emily.popgen.dk): "/course/advBinf/joel/Exercise".

First, you will estimate the birth-and-death rate and conduct evolutionary inferences of the Odorant Binding protein gene family in 18 Hexapoda species (Vieira and Rozas 2011) using the CAFE package. Using as input the table with OBP gene counts "obp_all.hexapoda.tsv" and the phylogenetic tree "hexapoda.nwk", answer the following questions:

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

Which is the general mode of evolution of the OBP gene family across Hexapoda? Has it undergone different dynamics across the surveyed species (i.e.: Is the OBP gene family expanded or contracted in any specific lineage)?

Now, you can explore if the birth-and-death rate would differ under a two-model lambda with different rates for *Drosophila* and non-*Drosophila* hexapods. Using the previous files, in addition to the newick tree with the 2-lambda model "hexapoda_2lambdamodel.nwk" you would be able to shed light into OBP evolution [Note: as an exploratory analysis, you only need to run the model and see the results, but there is no need to test for signification by using simulations]:

Are there differences in the OBP dynamics between *Drosophila* and the other Hexapoda species surveyed? Is the birth-and-death process constant across the phylogeny?

We have learnt that the OBP gene family follows a birth-and-death model of evolution where, in summary, new copies arise by tandem duplication and they can gain new functions or be lost by pseudogenization. Now you will explore the functional constraints in new duplicated copies by using a codon alignment of five *Drosophila melanogaster* copies (OBP56f, OBP56g and OBP56h are duplicated genes, OBP19a and OBP76b are used as outgroup sequences). Using HyPhy absrel and meme models in the "Obp_aln_absrel.fas [Used for absrel] and Obp_aln_meme.fas [Used for meme]" and "Obp_tree.nwk [Used for both absrel and meme]" files, you should be able to study the presence of functional constraints and positive selection in the codon alignments:

Which is the d_N/d_S ratio in the three duplicated OBP genes (OBP56f, OBP56g and OBP56h)? Are there different functional constraints among these copies? Which is the most likely reason of this observation?

Is there any specific site under positive selection?

Bibliography

Hahn MW, Han MV, Han SG (2007) Gene family evolution across 12 Drosophila genomes. PLoS Genet 3: e197.

Han, MV, Thomas, GWC, Lugo-Martinez, J, Hahn, MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol 30, 8.

Murrell, B. et al. (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genet 8, e1002764.

Smith, MD et al. (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol 32, 1342–1353.

Vieira FG, Rozas J (2011) Comparative Genomics of the Odorant-Binding and Chemosensory Protein Gene Families across the Arthropoda: Origin and Evolutionary History of the Chemosensory System. Genome Biol Evol 3: 476-490.