

UPPSALA UNIVERSITY
MOLECULAR EVOLUTION (1MB461)

Project

Author:
Daniel ÅGSTRAND



UPPSALA
UNIVERSITET

MARCH 7, 2020

1 Introduction

This projects goal is to look for positive selection in the OR2W1 gene in chimpanzees[1]. The OR2W1 codes for the protein Olfactory Receptor 2W1 which is a G-protein-coupled receptors (GPCR) in the large protein superfamily Olfactory Receptor which detect odorant molecules which give rise to a sense of smell. The Olfactory Receptor superfamily is a multigene family with over 800 genes in the chimpanzee genome, although only 47.85 is functional genes while the rest are pseudogenes or truncated genes[2]. And this is a low amount of Olfactory Receptor genes compared to other vertebrate.[3] This specific Olfactory Receptor has evolved to detect benzyl acetate[4] which is an organic ester that can be found in for instance in jasmine, apples and pears and humans perceives it as sweet and pleasant which is why it's often used in hygiene and health care products[5].

Because the Olfactory Receptor genes is a really diverge gene family and also because it would be good for a individual to be able to recognize as many important chemicals as possible, it's highly likely that the OR2W1 is under positive selection.

2 Method

The first thing that was done was to identify orthologs for the OR2W1 gene in chimpanzees. This was preformed with help from the Ensembl Genome Browser (<https://www.ensembl.org/>). After the orthologs had been identified, they were aligned with the help from the alignment tool on the Ensembl website. From the aligned sequence file a tree file was created with Maximum-Likelihood Phylogenies with the online tool PhyML 3.0 (<http://www.atgc-montpellier.fr/phyml/>).

The data was then analysed to be able to determine if the OR2W1 gene is under positive selection in chimpanzees. The analysis was preformed with different tests to check for the differences in non-synonymous and synonymous mutations between the specific gene and its orthologs. According to $\omega = dn/ds$, where dn is non-synonymous changes per non-synonymous site and ds is synonymous changes per synonymous site. This is because the ratio between those mutations can show if a specific gene is under positive, negative or neutral selection. If $\omega > 1$ then the gene is under positive selection, if $\omega < 1$ then it's under negative selection and if $\omega = 1$ then it's under neutral selection.

Three different complexities of models were used to preform the analysis. The simplest model that has been used assumes that the gene and the different orthologs is under the same selection pressure over all selected species and without dividing the the result into different sites for each specific selections. A more advanced model calculated different selection pressure for each selected species. The most complex model calculated different selection pressures for each species and also divided the sites into different selection pressures. Then a likelihood ratio test was preformed on each model to see how well the data fitted on that specific model.

3 Result

Sequence data

Chimpanzee (ENSPTRP00000030497), Human (ENSP00000366380), Gibbon (ENSNLEP00000023049), Dog (ENSCAFP00000033838), Alpaca (ENSVPAP00000007769), American beaver (ENSCCNP00000015207) and Hedgehog (ENSEEUP00000005787).

Test results and likelihood ratio test

The branch test gave a result of $\omega = 1.22932$ for the gene in chimpanzee and the p-score compared to the simplest model was 0.0034. The branch-site test gave the following result with chimpanzee as foreground and a p-score of 0.8049 when compared to a branch-site model with fixed ω :

| site class | 0 | 1 | 2a | 2b |
|--------------|---------|---------|---------|---------|
| proportion | 0.38584 | 0.10339 | 0.40283 | 0.10794 |
| background w | 0.09850 | 1.00000 | 0.09850 | 1.00000 |
| foreground w | 0.09850 | 1.00000 | 2.12916 | 2.12916 |

4 Discussion and Conclusion

The branch test shows positive selection for the OR2W1 gene in chimpanzees with data that fits well on the model. The branch-site test also shows positive selection but with results that really overfits the model. This could be because the branch-site model gets underpowered so even if there is positive selection in the OR2W1 gene the result isn't significant enough or because that the species are too closely related to each other so that the test might mistake some mutations for fixations because they are all identical in all species. But in this case this isn't likely because another group of orthologs from only primates has also been analysed to check if that would overfit the data even more (because those orthologs are even more closely related). But this wasn't the case, the orthologs from only primates fitted the model slightly better (it got a p-score of around 0.6504). Another reason could also be that the result from the branch test isn't statistically significant, which would mean that the branch-site test gets overfitted because there isn't positive selection on the OR2W1 gene in chimpanzees.

5 Referenser

- [1] Ensembl. ENSPTRG00000017886. Available on https://www.ensembl.org/Pan_troglodytes/Gene/Summary?db=core;g=ENSPTRG00000017886;r=6:29284994-29285956;t=ENSPTRT00000033013.
- [2] Niimura, Y. (2009). Evolutionary dynamics of olfactory receptor genes in chordates: interaction between environments and genomic contents. *Hum Genomics* 4, 107. Available on <https://doi.org/10.1186/1479-7364-4-2-107>.
- [3] Niimura, Y. Nei, M. (2007). Extensive Gains and Losses of Olfactory Receptor Genes in Mammalian Evolution. *PLOS ONE* 2(8): e708. Available on <https://doi.org/10.1371/journal.pone.0000708>.
- [4] Audouze, K., Tromelin, A., Le Bon, AM., Belloir, C. and Petersen, RK. (2014). Odorant-Receptor Interactions by Global Mapping of the Human Odorome. *PLOS ONE* 9(4): e93037. Available on <https://doi.org/10.1371/journal.pone.0093037>.
- [5] PubChem Database. Odorant-Receptor Interactions by Global Mapping of the Human Odorome. Available on <https://pubchem.ncbi.nlm.nih.gov/compound/Benzyl-acetate>.