Chapter 5 – The evolution of the retinol metabolism and its role in the origin of vision

Introduction

- What is the retinol metabolism and which functions/processes is it involved in

- Focus on role for visual function

- description of kegg pathway (Figure 1), roles of main components (should I know what each component does? Even if in intro I can be more generic)

- few words to highlight how some components are more specific to this pathway and others are more generic (Table 1). Also, specify how based on current knowledge the genes listed by kegg pathway can be grouped into 14 known gene families (Table 4.1), but we want to know if these families are consistent from an evolutionary perspective.

- Problematic/aims: from an evolutionary point of view, we need to check a) relationship amongst all these components (are some belonging to same orthogroup etc); b) once OGs are defined, understand their distribution across animals and eukaryotes; c) finally describe the major evolutionary events that characterise each OG.

Results & Discussion

- choice of components to study and their description (Refer to figure 1 and table 1)

- Show the comparison of OOG and BOG, highlighting how it compares to original kegg groups: can show both the “network” from cytoscape (Figure 4.2) and presence/absence tables with different versions (as a supplementary excel file) + a main table showing the final 12 OGs chosen for the phylogenetic analysis: a simple table with just name of OG/short name and number of sequences.

- Describe what is the final list of OGs. Linked to the p/a tables above.

- comparison of generax and possvm results for each OG: could be a figure in which on the right there is the annotated generax tree (events.newick) with the D & L nodes highlighted and main groups within the OG colour coded; on the left the same style but with the Possvm output. So the two methodologies can be easily compared: are there discrepancies for the groupings and/or events distributions?

“where applicable the possvm OG of interest is indicated” (in case of cyp multiple OGs..)

Interesting to check number of possvm OG / num seqs and compare with num generx events / num seqs. The former may indicate fragmentation within OG, complex substructure. The latter shows that OG is evolving quickly (?) as multiple events are occurring. In theory we expect them to overlap/follow similar trends? (sure?)

- Also give for each OG a summary of number of events as measure of amount of evolutionary change that has been going on in the OGs. (I know how to retrieve this data from generax, does possvm have the same? If so compare).

Methods

- Definition of species to look into

- Collection of kegg “orthologs” for each components as queries

- preliminary loose blastp

- definition of orthogroups: used to tools/methodologies: Orthofinder and Broccoli

- To annotate OGs: eggnog, keep OGs with kegg map00830, see what human sequences in OG are annotated as.

- To annotate OGs: focus on what the human, and a handful of other species, genes were for each orthogroup.

- to annotate sequences: swissprot plus other species-specific databases.

- Comparison of OOG and BOG

- Using combined results for further phylogenies

- Eggnog annotation and highlighting of EggNog OGs : see how many sequences are similar to the hsap one...

- Reconstructing evolutionary events in each OG with two methodologies: generax and possvm

- Comparison of the two methodologies and description of events

Note: the idea is that with Possvm we can in theory do the reconstruction without a species tree. For generax we need to provide a species tree. We provide the sponge-first version (we can cite previous phototr chapter in which we had compared number of events between cteno-first and sponge-first and had found slightly less events (better/more parsimonious) in sponge-first scenario. And over all not huge difference overall between the two so it’s almost irrelevant which one we chose. But anyway however, just in case the choice of species tree given does bias results, we used the alternative methodology Possvm. However, Possvm also has option of giving starting species-tree. Which we tried. Now if the one with species tree is much better results, we should use that and not mention that we used Possvm to avoid issue of species tree and just said we compared two different approaches. If version without species tree is really good, we could use that.

Also note: the filtering by Eggnog OGs is very biased towards human type of OGs, even if we do Possvm with also filtered dataset, it’s probably best to use the non-filtered one. Also because then one might ask why we didn’t do also generax both with and without. Perhaps it’s best to keep the eggnog “filtering” just as extra supplementary info, to say look regarding what we know about human genes, we can define these groups, from our dataset x,y,z sequences fall in that category.

Conclusions

- Discuss the use of comparing different methodologies that can be used to have bigger picture and are complementary to each other etc.

- Focus a little bit on orthofinder vs broccoli

- Focus a little bit of generax vs possvm

- Describe what we now know about the evolution and distribution of the different OGs and implications for the evolution of retinol metabolism as a whole. Focus on implications for vision. This is where it is useful to know which components are essential for the 11-cis retinal formation. If some components are missing are they essential? Could their function potentially be done by unrelated molecules not currently associated to the metabolism pathway.

Future perspectives: e.g. identifying potential alternative pathways that could be working in other species? Mass spec? Choose carefully what to talk about!!

LIST OF MATERIAL AND WHERE TO FIND IT:

- Powerpoint with kegg components: all; which present in a selection of species; a colour coded version based on orthofinder OGs (note: this should be updated to colour code based on OOG+BOG): Uni Leicester Onedrive: C:\Users\ale\_a\OneDrive - University of Leicester\Retinol\_Project\Retinol\_Met\_maps

- List of components and info related to them (e.g. how many pathways involved in according to kegg) – both on uni onedrive, but perhaps more updated version on google drive

- comparison of presence absence based on kegg, OOG, BOG – google drive

- network of OOG vs BOGs (both how comparison was done and the figure) – ALICE (probably on /data but some parts maybe already in /rfs)