General Discussion, Conclusions, and Future Perspectives

Note: some generic words about how the projects presented in this thesis are large scale exploratory works that have also the objective to direct future targeted investigations. Targeted both in the sense of topic (specific questions that my projects have opened up; e.g. TAFA in urochordates... (an example for each chapter).. ) but also in a methodological sense (can be approached with specific experimental approaches – since they are expensive and time consuming it is good to have this type of exploratory large scale bioinformatics works first).

Comparison of methodologies to define orthogroups/gene families: family-specific strategies (e.g. as used in chapter 3) versus blind/broadscale approaches (e.g. chapter 4). Chapter 5 is not very comparable as there was only few families to look at so makes sense to have a targeted approach. The problem is when you want to look at contemporarily multiple different families and each very different/unrelated. The strategy of doing family-specific data mining (ch 3) is more precise and gives you more control over the process without the risk of “missing” errors. It also allows you to decide the breadth of analyses you want to do, e.g. whole family or subset of family? (e.g. all g alpha proteins or only G alpha i/o type). However, it is very time consuming. The approach of using algorithms that detect orthogroups (broccoli, orthofinder) is very quick and powerful. They can be used for very large datasets. Although there is less control over the findings, and using mini proteomes (like I did, in which species proteomes had been filtered out with a preliminary loose blast) may cause some biases. Ultimately, though since these datasets are then used for rigorous phylogenetic analysis, any incongruounses within the orthogroups may be identified (here I might have some examples to discuss e.g. what is situation in SDR+RDH?).