Chapter 6

General Discussion, Conclusions and Future Perspectives

General Discussion

The power of evolutionary studies in understanding fundamental animal processes

Fundamental aspects of animal biology are deeply rooted in evolutionary history, especially the formative stages characterised by the major transition from a unicellular ancestor to a state of obligate multicellularity. The changes in genetic make-up that our early animal ancestors must have undergone are inextricably tied to this profound shift in lifestyle. The subdivision of specific tasks amongst different cells resulted in the origin of different cell types. Contemporarily, the need for cells of a multicellular organism to communicate and coordinate with each other both for internal physiology and for responding collectively to external stimuli, presented a critical challenge (Ruiz-Trillo and Nedelcu 2015). Such evolutionary pressures drove the expansion of gene families involved in cell signalling (Paps and Holland 2018). The evolution of vision, for instance, required a specialised cell type, the photoreceptor cell, capable both of detecting light, and of relaying this signal to other cells, that in turn can integrate this signal to elaborate a collective response. This was possible by coupling photosensitive molecules, the opsin bound to a cis-retinal molecule, with a phototransduction machinery capable of transducing the signal within the cell and culminating in ion channel modulation. The resulting flow of ions is responsible for the electrical signalling that starts the communication with other cells (Hardie and Juusola 2015; Lamb 2020). Similarly, cells of the immune system are responsible for maintaining the physiology of an organism both under normal conditions and when faced with external threats. Chemokine signalling, for example, plays an essential role in numerous processes in vertebrates by coordinating cell migrations throughout the body during development, homeostasis, and host defence. This system signals through small chemotactic proteins secreted by certain cells that are recognised by GPCR receptors on the target cells, which are induced to move along the gradient of these molecules (Murphy 2023).

Acknowledging this evolutionary history and recognising the centrality of cell signalling for the maintenance of multicellularity is essential for a deeper understanding of animal biology and evolution. This principle was an inspiration for the research presented in this thesis, where I explored the evolution of vision and the chemokine system as examples of processes relying on cell signalling. To explore the evolutionary history of vision two primary aims were identified: first, investigating the evolution of photoreceptor cells by tracing the history of the phototransduction pathway components and the regulatory genes involved in the cell type specification; and second, reconstructing the evolution of the retinol metabolism pathway, which is responsible for the constant production of the photosensitive retinal that, when bound to opsin, enables the phototransduction process. To investigate the evolution of chemokine signalling, three main aims were pursued: first, clarifying the relationships between the “canonical” chemokines and chemokine receptors and “non-canonical” proteins that are also involved in the system; second, reconstructing the evolution of the ligand components; lastly, reconstructing the evolution of the receptor components. To address these aims, a combination of large-scale phylogenetic and bioinformatic methods were employed. Instead of limiting the focus to only a subset of molecular components, the analyses were conducted on a comprehensive set of components from the pathways and systems under investigation. This approach allowed to collect multiple pieces of evidence that combined created a broader picture of the trajectories of the systems under study, offering new insights into their evolution.

Animal-specific gene expansions as the molecular foundation of vision

The phylogenetic analyses of phototransduction gene families suggest an ancient origin in most cases. Many of these expansive families are present across Holozoa, with several of them being widely distributed throughout all eukaryotes. The involvement of these broad gene families in a variety of different pathways and processes other than vision, can explain their presence outside of animals. The few exceptions are opsins and G gamma, integral to both ciliary and rhabdomeric type phototransduction, along with the inhibitory subunit of the PDE6 enzyme and RGS9BP, specific to ciliary phototransduction. These families appear to be exclusive to animals, and in some cases to vertebrates. Within the extensive gene families, distinct subfamilies can be delineated. This granularity provides enhanced clarity as it enables a more detailed tracing of the evolutionary paths of those subfamilies containing genes known to be integral to phototransduction pathways in model organisms. A recurring observation, for instance, is that while the overarching gene family may span eukaryotes, numerous expansions leading to diverse subfamilies predominantly took place within holozoans, just before the emergence of animals, or within the early history of animals. It is within these more recent subfamilies that we find the exact genes, that are well-documented in their roles in phototransduction of model organisms such as human and flies. While these observations generally apply to gene families of both major phototransduction pathways—ciliary and rhabdomeric—some differences emerged. For instance, within certain ciliary phototransduction families, the subfamilies most closely associated to vision are vertebrate-specific. This occurs for example in the super family of the PDE6 catalytic subunit, where the PDE6A/B/C is present only in vertebrates, and within the RGS super family, where RGS9 is present only in vertebrates. This, combined with the fact that the inhibitory subunit of PDE6 (PDE6G/H) appears to be vertebrate-specific implies that while foundational components of the ciliary pathway existed within all animals, some novelties, including expansions within older gene families and, possibly, *de novo* emergence of new genes, occurred concurrently with the evolution of vertebrates. While other members of the catalytic PDE6 subunit and RGS superfamilies likely can fulfil similar roles in non-vertebrates, the absence of PDE6G/H outside vertebrates suggests an auxiliary role. Indeed, in vertebrates this gene is involved in the shut-off step of phototransduction, contributing to improve the regulation and efficiency of the system in recovering from light stimuli (Lamb et al. 2018), rather than being essential for the basic phototransduction response.

In addition to offering valuable insights into when the full suite of phototransduction components coalesced in the genome of our ancestors, the exhaustive search for phototransduction genes across a diverse range of organisms enabled the compilation of specific markers for photoreceptor cells. These markers were subsequently used to detect photoreceptor cell profiles in the single-cell data of various animals, notably pinpointing potential photoreceptor candidates in all four non-bilaterian phyla. However, not all phototransduction families were consistently present in candidate photoreceptor cells of non-bilaterian species. Moreover, distinguishing between the prevalence of ciliary or rhabdomeric pathways proved challenging. This suggests that these phyla might not strictly adhere to the classical ciliary or rhabdomeric pathways. Instead, they could exhibit unique, lineage-specific variations. This, to some extent, seemed to be the case also for some bilaterian organisms such as sea squirt and sea urchin. Our current understanding of phototransduction is heavily influenced by a surprisingly narrow selection of organisms. As such, it is plausible that these do not fully represent the true diversity of phototransduction systems across the animal kingdom. Research on the evolution of vision has, in recent years, seen an increase of interest in examining early-branching animals (Sebé-Pedrós, Saudemont, et al. 2018; Sebé-Pedrós, Chomsky, et al. 2018; Levy et al. 2021; Wong et al. 2022; McCulloch et al. 2023) —just like this study—as they offer key insights into the ancestral state of photoreceptor cells and vision. However, I argue that it is just as crucial not to overlook other bilaterian species. These could potentially serve as intermediaries, bridging our understanding between distantly related non-bilaterians and the handful of well-studied model organisms. It is only with a comprehensive grasp of the diversity of animal phototransduction pathways that it will be possible to fully trace the patterns that characterise their evolution. We are seeing exciting times, in which single-cell datasets of more and more animal species are being published (Gavriouchkina et al. 2022; Lust et al. 2022; Babonis et al. 2023; McCulloch et al. 2023; Piovani et al. 2023; Tominaga et al. 2023). Determining the evolutionary relationships among genes involved in phototransduction and pinpointing their presence or absence across a diverse spectrum of animals, is the first step for then exploring photoreceptor cell profiles in single cell data. This, combined with experimental studies, is contributing to further illuminating the intricate puzzle of vision evolution—a subject that has captivated researchers for centuries.

Similarly, a deeper understanding of the regulatory toolkits that characterise photoreceptor cells will also benefit from the availability of more and diverse single-cell datasets. However, interesting preliminary patterns are already apparent with the species that I was able to investigate in this thesis. Several families of regulatory genes frequently appeared in photoreceptor cells across animal species. These included, but were not limited to, transcription factors recognized for specifying the photoreceptor cell type in model organisms, like Six/3/6, Meis2, and Tbx2 in humans, and the ATF4 ortholog crc and glass in flies. While, the exact combinations of these regulatory genes were not conserved throughout animals, certain families of transcription factors were more frequent than others, such as bZIP transcription factor, zinc finger C2H2 and homeobox families. This observation challenges the approach classically used of focusing on orthologous transcription factors as markers of shared photoreceptor cell profiles (Arendt 2003). Instead, a broader lens that focuses on transcription factor families might be more appropriate, especially given the challenges of identifying orthologs in distantly related species. This alternative approach will benefit from large-scale phylogenetic analyses as well. Beyond transcription factors, which represent the most abundant type of regulatory genes shared across animal photoreceptor cells, other regulatory genes, including transcription cofactors and genes involved in chromatin conformation, were also consistently observed.

Finally, the last pieces of the puzzle that this thesis has to offer for understanding the evolution of vision, come from the phylogenetic analysis of the retinol metabolism enzymes that ensure the constant availability of cis-retinal, which bound to the opsin, is the first respondent to light stimuli. With this analysis, the retinol metabolism enzymes were found to belong to 12 distinct orthogroups—phylogenetically defined gene families. This confirmed some established family relationships but also revealed some surprises, such as the separation of the Diacylglycerol O-Acyltransferase (DGAT) enzymes into two distinct orthogroups. Overall, all orthogroups were revealed to have ancient origin and widespread distribution across eukaryotes. The only exception was the Lecithin Retinol Acyltransferase (LRAT) orthogroup, which exhibited a sporadic distribution both within and outside animals. Further refined phylogenetic analyses for each orthogroup yielded deeper insights into the substructures of these broad gene families as well as their distribution. An intriguing observation was that enzyme families with a marginal role in the pathway often had subgroups distributed widely across eukaryotes. Conversely, the subfamilies of the most specific enzyme families, such as RPE65, which catalyses the hydrolysis of stored all-trans-retinyl esters to 11-cis retinol, were generally exclusive to animals. Ultimately, the large-scale phylogenetic approach employed to reconstruct the evolution of both the phototransduction pathways, and the retinol metabolism unveiled a common narrative. It suggests that broad gene families, which might have originally played roles in foundational biological processes, experienced lineage-specific expansions within animals. Some of the emergent subfamilies were then co-opted into roles vital for vision, which, without them, would not have achieved its present-day complexity and sophistication.

Evolutionary dynamics of chemokine signalling

A similar approach to that used to investigate vision, was also applied to investigate the evolution of the chemokine signalling system. Though the system fundamentally consists of two categories of components—ligands and receptors—it is complicated by the inclusion of “non-canonical” components. Additionally, chemokines, and to some extent their receptors, are small fast evolving molecules that have promiscuous interactions, thereby confusing their evolution. Employing sequence clustering techniques, my colleagues and I started by untangling the intricate relationships amongst canonical and non-canonical components. It was clarified that the non-canonical ligand families were not only distinct from canonical ones but also unrelated amongst themselves. In contrast, only one receptor group, the atypical chemokine receptor 1, was unrelated to all other receptors. These distinctions have important implications, as the literature has at times grouped disparate families based on their analogous functions and superficial sequence similarities (Tom Tang et al. 2004; Pisabarro et al. 2006; Weinstein et al. 2006; Tomczak and Pisabarro 2011; Chen et al. 2018; Liu et al. 2018; Zhang et al. 2018). While such categorizations may hold relevance in certain immunological contexts, they blur the true evolutionary relationships of these families. Furthermore, accurately discerning the relationships among these protein families is particularly important given their general pharmacological relevance and their key role as therapeutic targets (Lai and Mueller 2021).

Once the distinct groups of ligands and receptors were clarified, it was possible to examine their distribution across animals and investigate the details of each of their evolutionary histories. The strength of our research was the comprehensive examination of this system across a broad spectrum of animal species. While all canonical components were confirmed to be vertebrate-specific, several non-canonical components were found to be more ancient. For instance, TAFA ligands were detected also in urochordates, the sister group of vertebrates, and the CKLF super family was found across bilateria. This opens interesting avenues of future research in investigating the physiological roles of these molecules in invertebrates in the broader context of the evolution of immune systems. Finally, our detailed description of the pattern of duplication and loss events occurring within each gene family, provided an explanation for the diversity of the system. This ranged from uncovering ancient events, like the initial duplication at the stem of vertebrates which gave rise to all canonical and non-canonical receptors, to more recent occurrences such as the numerous mammalian-specific expansions observed in both chemokine and chemokine receptor groups.

Conclusions and Future Perspectives

By employing large-scale bioinformatic methods to investigate the origins and evolution of vision and chemokine signalling—two animal processes rooted in cell signalling— this thesis offered a multi-faceted perspective on the research questions. This approach led to various discoveries that can be further explored both bioinformatically and experimentally, laying foundation for future research in the field. Furthermore, this thesis provided an opportunity to experiment with various methodological strategies and bioinformatic tools tailored for expansive research questions, allowing for a deeper understanding of the advantages and drawbacks of each method. For instance, the curated strategy employed in Chapter 3 to identify the optimal protein profile for each individual gene family yielded highly accurate results. It also granted greater control over selecting gene families to analyse and offered flexibility in the scope of analyses, such as choosing between all GPCRs or focusing solely on opsins. Yet, this approach is also time-consuming, especially when investigating many and diverse gene families. Additionally, it necessitated a profound preliminary understanding of the targeted gene families to ensure informed decisions. On the other hand, utilizing algorithms that identify orthogroups, such those used in Chapter 4, is both rapid and robust. These tools are suitable for handling vast datasets. However, they offer less precision in determining the extent of the family, potentially including genes that might be too divergent from the original gene family of interest. The similarity-based clustering with CLANS used in Chapter 5 was in some respects an intermediary approach. While it was not as refined as the strategy of pinpointing family-specific profiles, and not as statistically robust as using orthogroup inferring tools, its adaptability in adjusting thresholds and intuitive visualization granted more control over determining the scope of the family collected. Ultimately, irrespective of the specific methods employed, for all gene families or orthogroups I conduct rigorous phylogenetic analyses, allowing me to detect any incongruences and place them in context. The lesson I learnt is that there is no single 'correct' method; it is instead essential is to consider the adequacy of the technique to the research question and the quantity and type of the gene families under study. To conclude, my comprehensive bioinformatic research on the evolution of the molecular components fundamental for vision and the chemokine system has provided a valuable starting point to direct future bioinformatic research and to select targeted questions to explore experimentally.

References

Arendt D. 2003. Evolution of eyes and photoreceptor cell types. *Int J Dev Biol* 47:563–571.

Babonis LS, Enjolras C, Reft AJ, Foster BM, Hugosson F, Ryan JF, Daly M, Martindale MQ. 2023. Single-cell atavism reveals an ancient mechanism of cell type diversification in a sea anemone. *Nat Commun* [Internet] 14:885. Available from: https://www.nature.com/articles/s41467-023-36615-9

Chen K, Bao Z, Tang P, Gong W, Yoshimura T, Wang JM. 2018. Chemokines in homeostasis and diseases. *Cell Mol Immunol* [Internet] 15:324–334. Available from: https://www.nature.com/articles/cmi2017134

Gavriouchkina D, Tan Y, Ziadi-Künzli F, Hasegawa Y, Piovani L, Zhang L, Sugimoto C, Luscombe N, Marlétaz F, Rokhsar DS. 2022. A single-cell atlas of bobtail squid visual and nervous system highlights molecular principles of convergent evolution. *bioRxiv* [Internet]:2022.05.26.490366. Available from: http://biorxiv.org/content/early/2022/05/28/2022.05.26.490366.abstract

Hardie RC, Juusola M. 2015. Phototransduction in Drosophila. *Current Opinion in Neurobiology* [Internet] 34:37–45. Available from: https://www.sciencedirect.com/science/article/pii/S0959438815000173

Lai WY, Mueller A. 2021. Latest update on chemokine receptors as therapeutic targets. *Biochemical Society Transactions* [Internet] 49:1385–1395. Available from: https://doi.org/10.1042/BST20201114

Lamb TD. 2020. Evolution of the genes mediating phototransduction in rod and cone photoreceptors. *Progress in Retinal and Eye Research* [Internet] 76:100823. Available from: https://www.sciencedirect.com/science/article/pii/S1350946219301107

Lamb TD, Patel HR, Chuah A, Hunt DM. 2018. Evolution of the shut-off steps of vertebrate phototransduction. *Open Biology* [Internet] 8:170232. Available from: https://royalsocietypublishing.org/doi/full/10.1098/rsob.170232

Levy S, Elek A, Grau-Bové X, Menéndez-Bravo S, Iglesias M, Tanay A, Mass T, Sebé-Pedrós A. 2021. A stony coral cell atlas illuminates the molecular and cellular basis of coral symbiosis, calcification, and immunity. *Cell* [Internet] 184:2973-2987.e18. Available from: https://www.sciencedirect.com/science/article/pii/S0092867421004402

Liu D-D, Song X-Y, Yang P-F, Ai Q-D, Wang Y-Y, Feng X-Y, He X, Chen N-H. 2018. Progress in pharmacological research of chemokine like factor 1 (CKLF1). *Cytokine* [Internet] 102:41–50. Available from: http://www.sciencedirect.com/science/article/pii/S1043466617303733

Lust K, Maynard A, Gomes T, Fleck JS, Camp JG, Tanaka EM, Treutlein B. 2022. Single-cell analyses of axolotl telencephalon organization, neurogenesis, and regeneration. *Science* [Internet] 377:eabp9262. Available from: https://www.science.org/doi/10.1126/science.abp9262

McCulloch KJ, Babonis LS, Liu A, Daly CM, Martindale MQ, Koenig KM. 2023. Nematostella vectensis exemplifies the exceptional expansion and diversity of opsins in the eyeless Hexacorallia. *EvoDevo* [Internet] 14:14. Available from: https://doi.org/10.1186/s13227-023-00218-8

Murphy PM. 2023. 15 - Chemokines and Chemokine Receptors. In: Rich RR, Fleisher TA, Schroeder HW, Weyand CM, Corry DB, Puck JM, editors. Clinical Immunology (Sixth Edition). New Delhi: Elsevier. p. 215–227. Available from: https://www.sciencedirect.com/science/article/pii/B9780702081651000150

Paps J, Holland PWH. 2018. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. *Nat Commun* [Internet] 9:1730. Available from: https://www.nature.com/articles/s41467-018-04136-5

Piovani L, Leite DJ, Guerra LAY, Simpson F, Musser JM, Salvador-Martínez I, Marlétaz F, Jékely G, Telford MJ. 2023. Single-cell atlases of two lophotrochozoan larvae highlight their complex evolutionary histories. :2023.01.04.522730. Available from: https://www.biorxiv.org/content/10.1101/2023.01.04.522730v2

Pisabarro MT, Leung B, Kwong M, Corpuz R, Frantz GD, Chiang N, Vandlen R, Diehl LJ, Skelton N, Kim HS, et al. 2006. Cutting Edge: Novel Human Dendritic Cell- and Monocyte-Attracting Chemokine-Like Protein Identified by Fold Recognition Methods. *The Journal of Immunology* [Internet] 176:2069–2073. Available from: https://www.jimmunol.org/content/176/4/2069

Ruiz-Trillo I, Nedelcu AM. 2015. Evolutionary Transitions to Multicellular Life: Principles and Mechanisms edited by Iñaki Ruiz-Trillo and Aurora M. Nedelcu. *Advances in Marine Genomics 2. Springer* [Internet] 91:370–371. Available from: https://www.journals.uchicago.edu/doi/abs/10.1086/688137

Sebé-Pedrós A, Chomsky E, Pang K, Lara-Astiaso D, Gaiti F, Mukamel Z, Amit I, Hejnol A, Degnan BM, Tanay A. 2018. Early metazoan cell type diversity and the evolution of multicellular gene regulation. *Nat Ecol Evol* [Internet] 2:1176–1188. Available from: https://www.nature.com/articles/s41559-018-0575-6

Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé M-P, Renno J, Loe-Mie Y, Lifshitz A, Mukamel Z, Schmutz S, et al. 2018. Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* [Internet] 173:1520-1534.e20. Available from: https://www.sciencedirect.com/science/article/pii/S0092867418305968

Tom Tang Y, Emtage P, Funk WD, Hu T, Arterburn M, Park EEJ, Rupp F. 2004. TAFA: a novel secreted family with conserved cysteine residues and restricted expression in the brain. *Genomics* [Internet] 83:727–734. Available from: https://www.sciencedirect.com/science/article/pii/S088875430300332X

Tomczak A, Pisabarro MT. 2011. Identification of CCR2-binding features in Cytl1 by a CCL2-like chemokine model. *Proteins: Structure, Function, and Bioinformatics* [Internet] 79:1277–1292. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.22963

Tominaga H, Nishitsuji K, Satoh N. 2023. A single-cell RNA-seq analysis of early larval cell-types of the starfish, Patiria pectinifera: Insights into evolution of the chordate body plan. *Developmental Biology* [Internet] 496:52–62. Available from: https://www.sciencedirect.com/science/article/pii/S0012160623000179

Weinstein EJ, Head R, Griggs DW, Sun D, Evans RJ, Swearingen ML, Westlin MM, Mazzarella R. 2006. VCC-1, a novel chemokine, promotes tumor growth. *Biochemical and Biophysical Research Communications* [Internet] 350:74–81. Available from: https://www.sciencedirect.com/science/article/pii/S0006291X06020122

Wong E, Anggono V, Williams SR, Degnan SM, Degnan BM. 2022. Phototransduction in a marine sponge provides insights into the origin of animal vision. *iScience* [Internet] 25:104436. Available from: https://www.sciencedirect.com/science/article/pii/S2589004222007076

Zhang K, Shi S, Han W. 2018. Research progress in cytokines with chemokine-like function. *Cellular & Molecular Immunology* [Internet] 15:660–662. Available from: https://doi.org/10.1038/cmi.2017.121